FORM PTO-1390 U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE	ATTORNEY'S DOCKET NUMBER
TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371	**PF-0727 USN**
	U.S. APPLICATION NO. (If known, see 37 CFR 1.5) **TO BE ASSIGNED** 10/049745

INTERNATIONAL APPLICATION NO.	INTERNATIONAL FILING DATE	PRIORITY DATE CLAIMED
PCT/US00/21878	**09 August 2000**	**09 August 1999**

TITLE OF INVENTION
PROTEASES AND PROTEASE INHIBITORS

APPLICANT(S) FOR DO/EO/US
YUE, Henry; LAL, Preeti; TANG, Y. Tom; BANDMAN, Olga; BAUGHN, Mariah R.; AZIMZAI, Yalda; LU, Dyung Aina M.; YANG, Junming

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is the **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. ☐ This is an express request to promptly begin national examination procedures (35 U.S.C. 371 (f)).
4. ☐ The US has been elected by the expiration of 19 months from the priority date (PCT Article 31).
5. ☒ A copy of the International Application as filed (35 U.S.C. 371(c)(2))
   a. ☐ is attached hereto (required only if not communicated by the International Bureau)
   b. ☐ has been communicated by the International Bureau.
   c. ☒ is not required, as the application was filed in the United States Receiving Office (RO/US).
6. ☐ An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)).
7. ☒ Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3))
   a. ☐ are attached hereto (required only if not communicated by the International Bureau).
   b. ☐ have been communicated by the International Bureau.
   c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
   d. ☐ have not been made and will not be made.
   e. ☐ attached hereto Article 34 Amendment
8. ☐ An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).
9. ☒ An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. ☐ An English language translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 11 to 16 below concern document(s) or information included:

11. ☐ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. ☒ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.27 and 3.31 is included.
13. ☒ A FIRST preliminary amendment, as follows: Cancel in this application original claims 12, 14, 18, 20, 21, 23, 24 & 27 before calculating the filing fee, without prejudice or disclaimer. Applicants submit that these claims were included in the application as filed in the interest of providing notice to the public of certain specific subject matter intended to be claimed, and are being canceled at this time in the interest of reducing filing costs. Applicants expressly state that these claims are not being canceled for reasons related to patentability, and are in fact fully supported by the specification as filed. Applicants expressly reserve the right to reinstate these claims or to add other claims during prosecution of this application or a continuation or divisional application. Applicants expressly do not disclaim the subject matter of any invention disclosed herein which is not set forth in the instantly filed claims.
    ☐ A SECOND or SUBSEQUENT preliminary amendment.
14. ☐ A substitute specification.
15. ☐ A change of power of attorney and/or address letter.
16. ☒ Other items or information:
1) Transmittal Letter (2 pp, in duplicate)
2) Return Postcard
3) Express Mail Label No.: **EL 856 146 595 US**
4) Sequence Listing Statement

JC13 Rec'd PCT/PTO 30 JAN 2002

U.S. APPLICATION NO. (if known, see 37 CFR 1.5) TO BE ASSIGNED 10/049745	INTERNATIONAL APPLICATION NO.: **PCT/US00/21878**	ATTORNEY'S DOCKET NUMBER **PF-0727 USN**

17. ☒ The following fees are submitted: **BASIC NATIONAL FEE (37 CFR 1.492(a)(1)-(5):** Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO.......$1000.00 ☐ International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO..$860.00 International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO........$710.00 ☒ **International preliminary examination fee paid to USPTO (37 CFR 1.482) but all claims did not satisfy provisions of PCT Article 33(1)-(4)..........$710.00** ☐ International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(1)-(4)..................$100.00		
ENTER APPROPRIATE BASIC FEE AMOUNT =	$710.00	
Surcharge of **$130.00** for furnishing the oath or declaration later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(e)).	$	

CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE		
Total Claims	20 =	0	**X $ 18.00**	$	
Independent Claims	2 =	0	**X $ 80.00**	$	
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			**+ $270.00**	$	
TOTAL OF ABOVE CALCULATIONS =				$	
☐ Applicant claims small entity status. See 37 CFR 1.27. The fees indicated above are reduced by 1/2.				$	
SUBTOTAL =				$710.00	
Processing fee of **$130.00** for furnishing the English translation later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1492(f)). +				$	
TOTAL NATIONAL FEE =				$710.00	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by the appropriate cover sheet (37 CFR 3.28, 3.31). **$40.00** per property +					
TOTAL FEES ENCLOSED =				$710.00	
				Amount to be Refunded:	$
				Charged:	$

a. ☐ A check in the amount of $________ to cover the above fees is enclosed.
b. ☒ Please charge my Deposit Account No. 09-0108 in the amount of $710.00 to cover the above fees.
c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 09-0108. A duplicate copy of this sheet is enclosed.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

INCYTE GENOMICS, INC.
3160 Porter Drive
Palo Alto, CA 94304

[signature]
SIGNATURE

NAME: Diana Hamlet-Cox

REGISTRATION NUMBER: 33,302

DATE: 20 January 2002

PROTEASES AND PROTEASE INHIBITORS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of proteases and protease inhibitors and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative and autoimmune/inflammatory disorders.

BACKGROUND OF THE INVENTION

Proteolytic processing is an essential component of normal cell growth, differentiation, remodeling, and homeostasis. The cleavage of peptide bonds within cells is necessary for the maturation of precursor proteins to their active forms, the removal of signal sequences from targeted proteins, the degradation of incorrectly folded proteins, and the controlled turnover of peptides within the cell. Proteases participate in apoptosis, inflammation, and tissue remodeling during embryonic development, wound healing, and normal growth. They are necessary components of bacterial, parasitic, and viral invasion and replication within a host. Four principal categories of mammalian proteases have been identified based on active site structure, mechanism of action, and overall three-dimensional structure. (See Beynon, R.J. and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York NY, pp. 1-5.)

The serine proteases (SPs) are a large family of proteolytic enzymes that include the digestive enzymes, trypsin and chymotrypsin; components of the complement cascade and of the blood-clotting cascade; and enzymes that control the degradation and turnover of macromolecules of the extracellular matrix. SPs are so named because of the presence of a serine residue found in the active catalytic site for protein cleavage. The active site of all SPs is composed of a triad of residues including the aforementioned serine, an aspartate, and a histidine residue. SPs have a wide range of substrate specificities and can be subdivided into subfamilies on the basis of these specificities. The main sub-families are trypases which cleave after arginine or lysine; aspases which cleave after aspartate; chymases which cleave after phenylalanine or leucine; metases which cleavage after methionine; and serases which cleave after serine. Clp protease is a unique member of the serine protease family as its activity is controlled by a regulatory subunit that binds and hydrolyzes ATP. Clp protease was originally found in plant chloroplasts but is believed to be widespread in both prokaryotic and eukaryotic cells (Maurizi, M.R. et al. (1990) J. Biol. Chem. 2665:12546-12552). SKD3, a mammalian homolog of the bacterial Clp regulatory subunit, has recently been identified in mouse (Perier, F. et al. (1995) Gene 152:157-163).

Cysteine proteases are involved in diverse cellular processes ranging from the processing of precursor proteins to intracellular degradation. Mammalian cysteine proteases include lysosomal

cathepsins and cytosolic calcium activated proteases, calpains. Of particular note, cysteine proteases are produced by monocytes, macrophages and other cells of the immune system which migrate to sites of inflammation and in their protective role secrete various molecules to repair damaged tissue. These cells may overproduce the same molecules and cause tissue destruction in certain disorders. In autoimmune diseases such as rheumatoid arthritis, the secretion of the cysteine protease, cathepsin C, degrades collagen, laminin, elastin and other structural proteins found in the extracellular matrix of bones. The cathepsin family of lysosomal proteases includes the cysteine proteases: cathepsins B, H, K, L, O2, and S; and the aspartyl proteases; cathepsins D and G. Various members of this endosomal protease family are differentially expressed. Some, such as cathepsin D, have a ubiquitous tissue distribution while others, such as cathepsin L, are found only in monocytes, macrophages, and other cells of the immune system.

Aspartic proteases include bacterial penicillopepsin, mammalian pepsin, renin, chymosin, and certain fungal proteases. The characteristic active site residues of aspartic proteases are a pair of aspartic acid residues, for example, Asp33 and Asp213 in penicillopepsin. Aspartic proteases are also called acid proteases because the optimum pH for their activity is between 2 and 3. In this pH range, one of the aspartate residues is ionized and the other is neutral. A potent inhibitor of aspartic proteases is the hexapeptide pepstatin which, in the transition state, resembles normal substrates.

Carboxypeptidases A and B are the principal mammalian representatives of the metallo-protease family. Both are exopeptidases of similar structure and active site configuration. Carboxypeptidase A, like chymotrypsin, prefers C-terminal aromatic and aliphatic side chains of hydrophobic nature, whereas carboxypeptidase B is directed toward basic arginine and lysine residues. Active site components include zinc, which coordinates two glutamic acid and one histidine residues in the protein.

Ubiquitin proteases are associated with the ubiquitin conjugation system (UCS), a major pathway for the degradation of cellular proteins in eukaryotic cells and some bacteria. The UCS mediates the elimination of abnormal proteins and regulates the half-lives of important regulatory proteins that control cellular processes such as gene transcription and cell cycle progression. In the UCS pathway, proteins targeted for degradation are conjugated to a ubiquitin, a small heat stable protein. The ubiquinated protein is then recognized and degraded by proteasome, a large, multisubunit proteolytic enzyme complex, and ubiquitin is released for reutilization by ubiquitin protease. The UCS is implicated in the degradation of mitotic cyclic kinases, oncoproteins, tumor suppressor genes such as p53, viral proteins, cell surface receptors associated with signal transduction, transcriptional regulators, and mutated or damaged proteins (Ciechanover, A. (1994) Cell 79:13-21). A murine proto-oncogene, Unp, encodes a nuclear ubiquitin protease whose overexpression leads to oncogenic transformation of NIH3T3 cells, and the human homolog of this gene is consistently elevated in small cell tumors and adenocarcinomas of the lung (Gray, D.A. (1995) Oncogene 10:2179-2183).

Protease inhibitors and other regulators of protease activity control the activity and effects of proteases. Protease inhibitors have been shown to control pathogenesis in animal models of proteolytic disorders (Murphy, G. (1991) Agents Actions Suppl. 35:69-76). Low levels of the cystatins, low molecular weight inhibitors of the cysteine proteases, correlate with malignant progression of tumors (Calkins, C. et al (1995) Biol. Biochem. Hoppe Seyler 376:71-80).

The plasma inter-α-trypsin inhibitor family molecules are serine protease inhibitors (serpins) composed of a 240 kDa plasma protein complex of at least five different types of glycoproteins. These glycoproteins consist of four heavy (H) chains and one 30 kDa light (L) chain named H1, H2, H3, H4, and L, and are independently synthesized and proteolytically processed from precursor proteins (Daveau, M. et al. (1998) Arch. Biochem. Biophys. 350:315-323; and Salier, J.P. et al. (1992) Mamm. Genome 2:233-239). The plasma inter-α-trypsin inhibitor light chains have sequence similarity to the Kunitz trypsin inhibitors which appear to be present in all vertebrates (Salier, J.P. (1990) Trends Biochem. Sci. 15:435-439). Some examples of the Kunitz trypsin inhibitors are tissue factor pathway inhibitor, which regulates tissue factor-induced coagulation, and protease nexin-2, which regulates serum coagulation factor XIa. (Broze, G.J. (1995) Annu. Rev. Med. 46:103-112; and Wagner, S.L. et al. (1993) Brain Res. 626:90-98). The heavy chain precursors encode a signal peptide sequence and the mature chain. Other plasma inter-α-trypsin inhibitor heavy chains have been described in human and rodents (Bourguignon, J. et al. (1993) Eur. J. Biochem.212:771-776; Salier, 1992, supra; and Salier, J.P. (1996) Biochem. J. 315:1-9). The expression of the rat plasma inter-α-trypsin inhibitor genes is regulated by inflammation in vivo. The genes are predominantly expressed in the rat liver, but H2 and H3 mRNA is also present in brain, intestine, and stomach (Daveau, supra.).

Kallistatins are members of the serine protease inhibitor family. Kallistatin forms a specific and covalently-linked complex with tissue kallikrein, which is a serine proteinase capable of cleaving kininogen to release vasoactive kinin. Components of the tissue kallikrein-kinin system include tissue kallikrein, kallistatin, kininogen, kinin, bradykininB1 and B2 receptors, and kininases (Chao, J. and L. Chao (1995) Biol. Chem. Hoppe Seyler 376:705-713).

Proteases and protease inhibitory molecules may contain amino acid sequence motifs which determine protein-protein interactions, such as the potential metal-binding site of von Willebrand factor type A3 (vWFA3) motif, glycine-amino acid-serine-amino acid-serine. This motif is also required for ligand interaction in the homologous I-type domains of integrins CR3 and LFA-1 (Huizinga, E.G. (1997) Structure 5:1147-1156).

Protease inhibitors play a major role in the regulation of the activity and effect of proteases. They have been shown to control pathogenesis in animal models of proteolytic disorders and in the treatment of HIV (Murphy, G. (1991) Agents Actions Suppl. 35:69-76; and Pakyz, A. and D. Israel (1997) J. Am. Pharm. Assoc. (Wash.) NS37:543-551).

The discovery of new proteases and protease inhibitors and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative and autoimmune/inflammatory disorders.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, proteases and protease inhibitors, referred to collectively as "PPIM" and individually as "PPIM-1," "PPIM-2," "PPIM-3," "PPIM-4," "PPIM-5," "PPIM-6," "PPIM-7," "PPIM-8," "PPIM-9," "PPIM-10," "PPIM-11," "PPIM-12," "PPIM-13," "PPIM-14," "PPIM-15," "PPIM-16," "PPIM-17," "PPIM-18," "PPIM-19," "PPIM-20," "PPIM-21," "PPIM-22," "PPIM-23," "PPIM-24," "PPIM-25," "PPIM-26," and "PPIM-27." In one aspect, the invention provides an isolated polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-27.

The invention further provides an isolated polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-27. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:28-54.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In

another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27.

The invention further provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions

whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional PPIM, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional PPIM,

comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional PPIM, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a

change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:28-54, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, ii) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, ii) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 shows polypeptide and nucleotide sequence identification numbers (SEQ ID NOs), clone identification numbers (clone IDs), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding PPIM.

Table 2 shows features of each polypeptide sequence, including potential motifs, homologous sequences, and methods, algorithms, and searchable databases used for analysis of PPIM.

Table 3 shows selected fragments of each nucleic acid sequence; the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis; diseases, disorders, or conditions associated with these tissues; and the vector into which each cDNA was cloned.

Table 4 describes the tissues used to construct the cDNA libraries from which cDNA clones encoding PPIM were isolated.

Table 5 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"PPIM" refers to the amino acid sequences of substantially purified PPIM obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of PPIM. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PPIM either by directly interacting with PPIM or by acting on components of the biological pathway in which PPIM participates.

An "allelic variant" is an alternative form of the gene encoding PPIM. Allelic variants may

result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding PPIM include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as PPIM or a polypeptide with at least one functional characteristic of PPIM. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding PPIM, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding PPIM. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent PPIM. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of PPIM is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of PPIM. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PPIM either by directly interacting with PPIM or by acting on components of the biological pathway in which PPIM participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')$_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind PPIM polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic PPIM, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution.

Compositions comprising polynucleotide sequences encoding PPIM or fragments of PPIM may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (PE Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the

side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

A "fragment" is a unique portion of PPIM or the polynucleotide encoding PPIM which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50% of a polypeptide) as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:28-54 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:28-54, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:28-54 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:28-54 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:28-54 and the region of SEQ ID NO:28-54 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-27 is encoded by a fragment of SEQ ID NO:28-54. A fragment of SEQ ID NO:1-27 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-27. For example, a fragment of SEQ ID NO:1-27 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-27. The precise length of a fragment of SEQ ID NO:1-27 and the region of SEQ ID NO:1-27 to which the fragment corresponds

are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full-length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full-length" polynucleotide sequence encodes a "full-length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (Apr-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for

nucleic acid hybridization are well known and can be found in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of PPIM which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of PPIM which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of PPIM. For example, modulation may

cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of PPIM.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an PPIM may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of PPIM.

"Probe" refers to nucleic acid sequences encoding PPIM, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example

Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding PPIM, or fragments thereof, or PPIM itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels

and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants, and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook, J. et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides

generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human proteases and protease inhibitors (PPIM), the polynucleotides encoding PPIM, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative and autoimmune/inflammatory disorders.

Table 1 lists the Incyte clones used to assemble full length nucleotide sequences encoding PPIM. Columns 1 and 2 show the sequence identification numbers (SEQ ID NOs) of the polypeptide and nucleotide sequences, respectively. Column 3 shows the clone IDs of the Incyte clones in which nucleic acids encoding each PPIM were identified, and column 4 shows the cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones and their corresponding cDNA libraries. Clones for which cDNA libraries are not indicated were derived from pooled cDNA libraries. In some cases, GenBank sequence identifiers are also shown in column 5. The Incyte clones and GenBank cDNA sequences, where indicated, in column 5 were used to assemble the consensus nucleotide sequence of each PPIM and are useful as fragments in hybridization technologies.

The columns of Table 2 show various properties of each of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3 shows potential phosphorylation sites; column 4 shows potential glycosylation sites; column 5 shows the amino acid residues comprising signature sequences and motifs; column 6 shows homologous sequences as identified by BLAST analysis; and column 7 shows analytical methods and in some cases, searchable databases to which the analytical methods were applied. The methods of column 7 were used to characterize each polypeptide through sequence homology and protein motifs.

The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions associated with nucleotide sequences encoding PPIM. The first column of Table 3 lists the nucleotide SEQ ID NOs. Column 2 lists fragments of the nucleotide sequences of column 1. These fragments are useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:28-54 and

to distinguish between SEQ ID NO:28-54 and related polynucleotide sequences. The polypeptides encoded by the selected fragments of SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, and SEQ ID NO:54 are useful, for example, as immunogenic peptides. Column 3 lists tissue categories which express PPIM as a fraction of total tissues expressing PPIM. Column 4 lists diseases, disorders, or conditions associated with those tissues expressing PPIM as a fraction of total tissues expressing PPIM. Column 5 lists the vectors used to subclone each cDNA library. Of particular note is the expression of SEQ ID NO:28 in gastrointestinal tissue. Of particular note is the tissue-specific expression of SEQ ID NO:51. Over 83% of the tissues expressing SEQ ID NO:51 are derived from gastrointestinal tissue, particularly the liver.

The columns of Table 4 show descriptions of the tissues used to construct the cDNA libraries from which cDNA clones encoding PPIM were isolated. Column 1 references the nucleotide SEQ ID NOs, column 2 shows the cDNA libraries from which these clones were isolated, and column 3 shows the tissue origins and other descriptive information relevant to the cDNA libraries in column 2.

SEQ ID NO:30 maps to chromosome 9 within the interval from 78.4 to 90.6 centiMorgans. This interval also contains a gene associated with cell proliferation.

SEQ ID NO:37 maps to chromosome 12 within the interval from 116.6 to 118.9 centiMorgans. This interval also contains a gene associated with a neurological disorder.

SEQ ID NO:47 maps to chromosome 4 within the interval from 99.2 to 105.2 centiMorgans. This interval also contains a gene associated with cardiovascular disease.

The invention also encompasses PPIM variants. A preferred PPIM variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the PPIM amino acid sequence, and which contains at least one functional or structural characteristic of PPIM.

The invention also encompasses polynucleotides which encode PPIM. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:28-54, which encodes PPIM. The polynucleotide sequences of SEQ ID NO:28-54, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding PPIM. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence

encoding PPIM. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:28-54 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:28-54. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of PPIM.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding PPIM, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring PPIM, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode PPIM and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring PPIM under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding PPIM or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding PPIM and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode PPIM and PPIM derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding PPIM or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:28-54 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA

polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (PE Biosystems, Foster City CA), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (PE Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (PE Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding PPIM may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library

does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, PE Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode PPIM may be cloned in recombinant DNA molecules that direct expression of PPIM, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express PPIM.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter PPIM-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of PPIM, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively,

fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding PPIM may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, PPIM itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (PE Biosystems). Additionally, the amino acid sequence of PPIM, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active PPIM, the nucleotide sequences encoding PPIM or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding PPIM. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding PPIM. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding PPIM and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding PPIM and appropriate transcriptional and translational control

elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding PPIM. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding PPIM. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding PPIM can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding PPIM into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of PPIM are needed, e.g. for the production of antibodies,

vectors which direct high level expression of PPIM may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of PPIM. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, supra; and Scorer, supra.)

Plant systems may also be used for expression of PPIM. Transcription of sequences encoding PPIM may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, supra; Broglie, supra; and Winter, supra.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding PPIM may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses PPIM in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of PPIM in cell lines is preferred. For example, sequences encoding PPIM can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue

culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *apr*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding PPIM is inserted within a marker gene sequence, transformed cells containing sequences encoding PPIM can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding PPIM under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding PPIM and that express PPIM may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of PPIM using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on PPIM is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al.

(1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding PPIM include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding PPIM, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding PPIM may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode PPIM may be designed to contain signal sequences which direct secretion of PPIM through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding PPIM may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric PPIM protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of PPIM activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such

moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the PPIM encoding sequence and the heterologous protein sequence, so that PPIM may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled PPIM may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

PPIM of the present invention or fragments thereof may be used to screen for compounds that specifically bind to PPIM. At least one and up to a plurality of test compounds may be screened for specific binding to PPIM. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of PPIM, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which PPIM binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express PPIM, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing PPIM or cell membrane fractions which contain PPIM are then contacted with a test compound and binding, stimulation, or inhibition of activity of either PPIM or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with PPIM, either in solution or affixed to a solid support, and detecting the binding of PPIM to the compound. Alternatively, the

assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

PPIM of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of PPIM. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for PPIM activity, wherein PPIM is combined with at least one test compound, and the activity of PPIM in the presence of a test compound is compared with the activity of PPIM in the absence of the test compound. A change in the activity of PPIM in the presence of the test compound is indicative of a compound that modulates the activity of PPIM. Alternatively, a test compound is combined with an in vitro or cell-free system comprising PPIM under conditions suitable for PPIM activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of PPIM may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding PPIM or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding PPIM may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding PPIM can also be used to create "knockin" humanized animals

(pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding PPIM is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress PPIM, e.g., by secreting PPIM in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of PPIM and proteases and protease inhibitors. In addition, the expression of PPIM is closely associated with cell proliferation, inflammation, the immune response, and gastrointestinal, neurological, and reproductive tissue. Therefore, PPIM appears to play a role in cell proliferative and autoimmune/inflammatory disorders. In the treatment of disorders associated with increased PPIM expression or activity, it is desirable to decrease the expression or activity of PPIM. In the treatment of disorders associated with decreased PPIM expression or activity, it is desirable to increase the expression or activity of PPIM.

Therefore, in one embodiment, PPIM or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PPIM. Examples of such disorders include, but are not limited to, a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic

anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma.

In another embodiment, a vector capable of expressing PPIM or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PPIM including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified PPIM in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PPIM including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of PPIM may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PPIM including, but not limited to, those listed above.

In a further embodiment, an antagonist of PPIM may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PPIM. Examples of such disorders include, but are not limited to, those cell proliferative and autoimmune/inflammatory disorders described above. In one aspect, an antibody which specifically binds PPIM may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express PPIM.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding PPIM may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PPIM including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of PPIM may be produced using methods which are generally known in the art. In particular, purified PPIM may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind PPIM. Antibodies to PPIM may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally

preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with PPIM or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to PPIM have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of PPIM amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to PPIM may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce PPIM-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for PPIM may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion

of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between PPIM and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering PPIM epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for PPIM. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of PPIM-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple PPIM epitopes, represents the average affinity, or avidity, of the antibodies for PPIM. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular PPIM epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the PPIM-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of PPIM, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of PPIM-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al., supra.)

In another embodiment of the invention, the polynucleotides encoding PPIM, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA,

PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding PPIM. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding PPIM. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding PPIM may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in PPIM expression or regulation causes disease, the expression of PPIM from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in PPIM are treated by constructing mammalian expression vectors encoding PPIM and introducing these vectors by mechanical means into PPIM-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of PPIM include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). PPIM may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding PPIM from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to PPIM expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding PPIM under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate

vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4+ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding PPIM to cells which have one or more genetic abnormalities with respect to the expression of PPIM. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544; and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding PPIM to target cells which have one or more genetic abnormalities with respect to the expression of PPIM. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing PPIM to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res.169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction

and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding PPIM to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full-length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for PPIM into the alphavirus genome in place of the capsid-coding region results in the production of a large number of PPIM-coding RNAs and the synthesis of high levels of PPIM in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of PPIM into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme

molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding PPIM.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding PPIM. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding PPIM. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased PPIM expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding PPIM may be therapeutically useful, and in the treament of disorders associated with decreased PPIM expression or activity, a compound which specifically promotes expression of the

polynucleotide encoding PPIM may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding PPIM is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding PPIM are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding PPIM. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which

generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of PPIM, antibodies to PPIM, and mimetics, agonists, antagonists, or inhibitors of PPIM.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising PPIM or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, PPIM or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example PPIM or fragments thereof, antibodies of PPIM, and agonists, antagonists or inhibitors of PPIM, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by

standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind PPIM may be used for the diagnosis of disorders characterized by expression of PPIM, or in assays to monitor patients being treated with PPIM or agonists, antagonists, or inhibitors of PPIM. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for PPIM include methods which utilize the antibody and a label to detect PPIM in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring PPIM, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of PPIM expression. Normal or standard values for PPIM expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibody to PPIM under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by

various methods, such as photometric means. Quantities of PPIM expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding PPIM may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of PPIM may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of PPIM, and to monitor regulation of PPIM levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding PPIM or closely related molecules may be used to identify nucleic acid sequences which encode PPIM. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding PPIM, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the PPIM encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:28-54 or from genomic sequences including promoters, enhancers, and introns of the PPIM gene.

Means for producing specific hybridization probes for DNAs encoding PPIM include the cloning of polynucleotide sequences encoding PPIM or PPIM derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding PPIM may be used for the diagnosis of disorders associated with expression of PPIM. Examples of such disorders include, but are not limited to, a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and an autoimmune/inflammatory

disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma. The polynucleotide sequences encoding PPIM may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered PPIM expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding PPIM may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding PPIM may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding PPIM in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of PPIM, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding PPIM, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the

presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding PPIM may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding PPIM, or a fragment of a polynucleotide complementary to the polynucleotide encoding PPIM, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding PPIM may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding PPIM are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San

Diego CA).

Methods which may also be used to quantify the expression of PPIM include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described in Seilhamer, J.J. et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, incorporated herein by reference. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, antibodies specific for PPIM, or PPIM or fragments thereof may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies,

or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-

dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for PPIM to quantify the levels of PPIM expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological

sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding PPIM may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, e.g., Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding PPIM on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal

associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, PPIM, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between PPIM and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with PPIM, or fragments thereof, and washed. Bound PPIM is then detected by methods well known in the art. Purified PPIM can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PPIM specifically compete with a test compound for binding PPIM. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PPIM.

In additional embodiments, the nucleotide sequences which encode PPIM may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, in particular U.S. Ser. No. 60/147,986 and U.S. Ser. No. 60/160,807, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), pcDNA2.1 plasmid (Invitrogen, Carlsbad CA), or pINCY plasmid (Incyte Genomics, Palo Alto CA). Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled

water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (PE Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (PE Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VI.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the tools, programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were generated using the default parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences

and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programing, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and PFAM to acquire annotation using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:28-54. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

IV. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{5 \times \text{minimum } \{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every

mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding PPIM occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation, trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table 3.

V. Chromosomal Mapping of PPIM Encoding Polynucleotides

The cDNA sequences which were used to assemble SEQ ID NO:28-54 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:28-54 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 5). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

The genetic map locations of SEQ ID NO:30, SEQ ID NO:37, and SEQ ID NO:47 are described in The Invention as ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. A disease associated with the public and Incyte sequences located within the indicated interval is also reported in the Invention.

VI. Extension of PPIM Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:28-54 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and β-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E.

coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems).

In like manner, the polynucleotide sequences of SEQ ID NO:28-54 are used to obtain 5' regulatory sequences using the procedure above, along with oligonucleotides designed for such extension, and an appropriate genomic library.

VII. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:28-54 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

VIII. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra), mechanical

microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/µl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/µl RNase inhibitor, 500 µM dATP, 500 µM dGTP, 500 µM dTTP, 40 µM dCTP, 40 µM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 µl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a

resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

IX. Complementary Polynucleotides

Sequences complementary to the PPIM-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring PPIM. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of PPIM. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is

designed to prevent ribosomal binding to the PPIM-encoding transcript.

X. Expression of PPIM

Expression and purification of PPIM is achieved using bacterial or virus-based expression systems. For expression of PPIM in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express PPIM upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of PPIM in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding PPIM by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, PPIM is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from PPIM at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified PPIM obtained by these methods can be used directly in the assays shown in Examples XI and XV.

XI. Demonstration of PPIM Activity

Protease activity of PPIM is measured by the hydrolysis of appropriate synthetic peptide substrates conjugated with various chromogenic molecules. The degree of hydrolysis is quantified by spectrophotometric (or fluorometric) absorption of the released chromophore (Beynon, R.J. and J.S.

Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York NY, pp.25-55). Peptide substrates are designed according to the category of protease activity as endopeptidase (serine, cysteine, aspartic proteases), animopeptidase (leucine aminopeptidase), or carboxypeptidase (Carboxypeptidase A and B, procollagen C-proteinase). Chromogens commonly used are 2-naphthylamine, 4-nitroaniline, and furylacrylic acid. Assays are performed at ambient temperature using an aliquot of PPIM and the appropriate substrate in a suitable buffer. Reactions are carried out in an optical cuvette and followed by the measurement of increase/decrease in absorbance of the chromogen released during hydrolysis of the peptide substrate. The change in absorbance is proportional to PPIM activity in the assay.

XII. Functional Assays

PPIM function is assessed by expressing the sequences encoding PPIM at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT plasmid (Life Technologies) and pCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of PPIM on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding PPIM and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads

coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding PPIM and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIII. Production of PPIM Specific Antibodies

PPIM substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the PPIM amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (PE Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-PPIM activity by, for example, binding the peptide or PPIM to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XIV. Purification of Naturally Occurring PPIM Using Specific Antibodies

Naturally occurring or recombinant PPIM is substantially purified by immunoaffinity chromatography using antibodies specific for PPIM. An immunoaffinity column is constructed by covalently coupling anti-PPIM antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing PPIM are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of PPIM (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/PPIM binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and PPIM is collected.

XV. Identification of Molecules Which Interact with PPIM

PPIM, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules

previously arrayed in the wells of a multi-well plate are incubated with the labeled PPIM, washed, and any wells with labeled PPIM complex are assayed. Data obtained using different concentrations of PPIM are used to calculate values for the number, affinity, and association of PPIM with the candidate molecules.

Alternatively, molecules interacting with PPIM are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989, Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

PPIM may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	28	088718	LIVRNOT01	088718H1(LIVRNOT01), 151754F1 (FIBRAGT01), 151754R1 (FIBRAGT01), SCEA00861V1, SCEA01403V1, SCEA03107V1, SCEA01683V1
2	29	114551	TESTNOT01	1273531F1 (TESTTUT02), 1498122F1 (SINTBST01), 1686926F6 (PROSNOT15), 1922870R6 (BRSTTUT01), 2270121R6 (UTRSNOT02), 3227104F6 (COTRNOT01)
3	30	1261376	SYNORAT05	428341R6 (BLADNOT01), 488402R6 (HNT2AGT01), 1261376H1 (SYNORAT05), 1261376T6 (SYNORAT05), 1413230F6 (BRAINOT12), 1448134F1 (PLACNOT02), 1869342F6 (SKINBIT01), 2263303H1 (UTRSNOT02), 2365444T6 (ADRENOT07), 2875019H1 (THYRNOT10), 2908347H1 (THYMNOT05), 3818352H1 (BONSTUT01), g3840298, g1965665, g848456
4	31	1299481	BRSTNOT07	1299481H1 (BRSTNOT07), 1302262F6 (PLACNOT02), 1596742X330D1 (BRAINOT14), 1725693F6 (PROSNOT14), 2125677X306D3 (BRSTNOT07), SCHA02258V1, SCHA00613V1, g1477302
5	32	1873139	LEUKNOT02	003818R1 (HMC1NOT01), 1873139F6 (LEUKNOT02), 1873139X325D1 (LEUKNOT02), 1873139X326V1 (LEUKNOT02), 1899870F6 (BLADTUT06), 2510118F6 (CONUTUT01)
6	33	1903112	OVARNOT07	1903112H1 (OVARNOT07), 1905330T6 (OVARNOT07), 2509325H1 (CONUTUT01), 2621121R6 (KERANOT02)
7	34	1993044	CORPNOT02	1858513F6 (PROSNOT18), 1993044H1 (CORPNOT02), 3733554F6 (SMCCNOS01), 4749046H1 (SMCRUNT01), 4960159H1 (TLYMNOT05), 5397428H1 (LIVRTUT13), SBCA07095F3
8	35	2292182	BRAINON01	2199554H1 (SPLNFET02), 2199554X305B1 (SPLNFET02), 2292182R6 (BRAINON01), 3480414T6 (OVARNOT11), 5427954H1 (THYMTUT03)
9	36	2331301	COLNNOT11	1253717H1 (LUNGFET03), 2331301H1 (COLNNOT11), 2331301R6 (COLNNOT11)
10	37	2517512	BRAITUT21	1222614R1 (COLNTUT02), 1486943F6 (UCMCL5T01), 1486943T6 (UCMCL5T01), 1569195F1 (UTRSNOT05), 1813007F6 (PROSTUT12), 2517512H1 (BRAITUT21), 5847584H1 (BRAENOT04)
11	38	3489039	EPIGNOT01	2541141F6 (BONRTUT01), 3489039H1 (EPIGNOT01), 4871852H1 (COLDNOT01)

Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
12	39	5432879	SPLNNOT17	1429082F6 (SINTBST01), 1807480F6 (SINTNOT13), 2303440H1 (BRSTNOT05), 2669584F6 (ESOGTUT02), 3073745H1 (BONEUNT01), 3190142R6 (THYMNON04), 4693457H2 (BRAENOT02), 4774453H1 (BRAQNOT01), 5432879H1 (SPLNNOT17), g836070
13	40	5853753	FIBAUNT02	834033T1 (PROSNOT07), 1521711F6 (BLADTUT04), 1757751R6 (PITUNOT03), 2161634F6 (ENDCNOT02), SAEA01666R1, SCGA11716V1, SCGA05971V1, SCGA07285V1
14	41	411344	BRSTNOT01	411344F1 (BRSTNOT01), 411344H1 (BRSTNOT01), 411344R1 (BRSTNOT01), 1859850F6 (PROSNOT18), 2183379F6 (SININOT01), 2474963H1 (SMCANOT01), 2546619X300D1 (UTRSNOT11), 3728811H1 (SMCCNON03), 3932959H1 (PROSTUT09)
15	42	1256390	MENITUT03	1256390H1 (MENITUT03), SBAA04311F1, SBAA04104F1, SBAA03263F1, SBAA01188F1
16	43	1786774	BRAINOT10	857246H1 (NGANNOT01), 1786774F6 (BRAINOT10), 1786774H1 (BRAINOT10), 1810671T6 (PROSTUT12), 5202653H1 (STOMNOT08)
17	44	1911808	CONNTUT01	1255942F6 (MENITUT03), 1354692F6 (LUNGNOT09), 1354692T1 (LUNGNOT09), 1418156T1 (KIDNNOT09), 1436123F6 (PANCNOT08), 1498302T1 (SINTBST01), 1735923X304D1 (COLNNOT22), 1735923X318D4 (COLNNOT22), 1834236R6 (BRAINON01), 1911808F6 (CONNTUT01), 1911808H1 (CONNTUT01), 2360308H1 (LUNGFET05), 3075823H1 (BONEUNT01), 4106766H1 (BRSTTUT17), 5713020H1 (MASTTXT01)
18	45	1973875	UCMCL5T01	1220149R6 (NEUTGMT01), 1377281F1 (LUNGNOT10), 1377281T1 (LUNGNOT10), 1508602F6 (LUNGNOT14), 1973875H1 (UCMCL5T01), 5098879F6 (EPIMNON05)
19	46	2323917	OVARNOT02	1609987F6 (COLNTUT06), 2012426R6 (TESTNOT03), 2012426T6 (TESTNOT03), 2323917H1 (OVARNOT02), 2323917T6 (OVARNOT02), 4851027H1 (TESTNOT10)
20	47	2754960	THP1AZS08	039061R6 (HUVENOB01), 580098H1 (BRAVTXT05), 2025465H1 (KERANOT02), 2754960H1 (THP1AZS08), 2754960R6 (THP1AZS08), 2754960X11F1 (THP1AZS08), 2754960X15F1 (THP1AZS08), 2754960X310U1 (THP1AZS08), 2754960X50F1 (THP1AZS08), 3821989T6 (BONSTUT01), g3736615

Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
21	48	3092341	BRSTNOT19	3092341H1 (BRSTNOT19), 3092341T6 (BRSTNOT19)
22	49	3658034	ENDPNOT02	2623516R6 (KERANOT02), 3658034F6 (ENDPNOT02), 3658034H1 (ENDPNOT02), 3658034T6 (ENDPNOT02), 5216522H1 (BRSTNOT35), 5590053H1 (ENDINOT02)
23	50	3883861	UTRSNOT05	858111H1 (NGANNOT01), 858233H1 (NGANNOT01), 1364808R1 (SCORNON02), 1861181F6 (PROSNOT19), 1906985T6 (OVARNOT07), 2687868H1 (LUNGNOT23), 2687868X366D1 (LUNGNOT23), 2721116X369D1 (LUNGTUT10), 3883861H1 (UTRSNOT05), 5217169H1 (BRSTNOT35)
24	51	4993873	LIVRTUT11	4987943H1 (LIVRTUT10), 4993873H1 (LIVRTUT11), SCEA01665V1, SCEA00232V1, SXBC01625V1, SXBC01802V1, SCSA03627V1
25	52	5208004	BRAFNOT02	4696870F6 (BRALNOT01), 5208004H1 (BRAFNOT02)
26	53	5267783	BRAFDIT02	220636R1 (STOMNOT01), 679457R6 (UTRSNOT02), 1330537F6 (PANCNOT07), 1808720F6 (PROSTUT12), 1969475H1 (BRSTNOT04), 2697426F6 (UTRSNOT12), 2991180H1 (KIDNFET02), 3532849H1 (KIDNNOT25), 4992376F6 (LIVRTUT11), 5004695F6 (PROSTUT21), 5267783H1 (BRAFDIT02)
27	54	5583922	FIBAUNT01	726878R1 (SYNOOAT01), 956818X11 (KIDNNOT05), 1658964X12 (URETTUT01), 1658964X13 (URETTUT01), 2544879F6 (UTRSNOT11), 3748858H1 (UTRSNOT18), 4761921H1 (PLACNOT05), 5043801H1 (PLACFER01)

Table 2

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences	Homologous Sequence	Analytical Methods and Databases
1	444	S91 T244 T251 S277 T386 T38 S182 T263 T373 Y346	N36 N180 N197 N295	Signal_peptide: M1-A23 Serpins (serine protease inhibitors): M1-P441, L68-L444	g1397241 RASP1	Motifs BLAST-GenBank HMMER SPScan HMMER-PFAM BLIMPS-BLOCKS ProfileScan BLAST_PRODOM BLAST_DOMO
2	565	S9 S19 T343 T458 S5 S58 S82 S114 S184 S185 S295 T382 T432 T476 T495 T543 S2 S5 S12 S25 S42 T169 S307 T337 S352 T357 T426 S513 T523 Y220 Y514	N112 N494	Ubiquitin carboxyl-terminal hydrolases family 2: G226-L243, Y235-I549	g2746775 Similar to peptidase family C19 (ubiquitin carboxyl-terminal peptidase)	Motifs BLAST-GenBank HMMER-PFAM BLIMPS-BLOCKS BLAST-PRODOM BLAST-DOMO
3	589	T43 S71 S181 S200 S260 S304 S312 T506 T572 T40 S66	N55 N126 N136 N164 N167 N302 N501	Ubiquitin family signature: M37-K107 Ubiquitin-associated domain: Q541-S586	g3873621 Similar to ubiquitin family	Motifs BLAST-GenBank HMMER-PFAM
4	775	T305 T2 S27 S43 S67 S392 S611 S615 T647 S665 S710 S729 S759 S96 T106 S217 S288 S301 S316 S432 S438 T443 S575 T719 S723 Y334	N49 N215 N322 N387 N468 N487 N497 N504 N508 N568 N600	Ubiquitin carboxyl-terminal hydrolases family 2: G112-L129, G193-L202, V230-C244, Y354-V391, N380-S401 Ubiquitin hydrolase carboxyl-terminal thiolesterase: G112-K211	g2739431 Hematopoietic-specific IL-2 deubiquitinating enzyme	Motifs BLAST-GenBank HMMER-PFAM BLIMPS-BLOCKS BLAST-PRODOM BLAST-DOMO

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences	Homologous Sequence	Analytical Methods and Databases
5	351	S9 S41 S48 S194 S201 T203 T257 S278 T322 T324 S129 S162 S181 S194 S225 T226 S348 Y271	N46 N123 N317	Ubiquitin carboxyl-terminal hydrolases family 2: L49-L337	g5410230 Ubiquitin-specific protease 3	Motifs BLAST-GenBank HMMER-PFAM BLIMPS-BLOCKS BLAST-PRODOM BLAST-DOMO
6	136	T30 S104 Y98		Dipeptidyl serine protease iv: I9-S128 Serine family prolyl endopeptidase: M4-I136	g577284 Dipeptidyl peptidase IV	Motifs BLAST-GenBank BLAST-PRODOM BLAST-DOMO
7	396	S24 S139 T168 T177 S198 S223 S279 T369 S26 S60 S223 S292	N166	Ubiquitin carboxyl-terminal hydrolase: E74-I283	g2854121 BRCA1 associated protein 1	Motifs BLAST-GenBank BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO HMMER-PFAM
8	246	S87 Y65	N94 N156 N195 N225	Zinc-binding metalloprotease domain: R121-H133		Motifs HMMER-PFAM
9	262	T32 S78 S85 T89 S125 S26 S170 S244	N168	Inter-alpha-trypsin glycoprotein inhibitor precursor: T32-T197		Motifs BLAST-PRODOM
10	406	S18 S37 T80 S98 S112 S178 T292 S298 T320 T391 S105 S212 S220 Y213	N14 N56 N176 N318		g3309170 COP9 complex subunit 4	Motifs BLAST-GenBank

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences	Homologous Sequence	Analytical Methods and Databases
11	172	T117 S135 S146		signal peptide motif: M1-G13 ATP-binding kinase: I6-E164 AAA-protein family: P4-M69	g3875433 Similar to ATP binding protein	Motifs BLAST-GenBank SPScan BLAST-PRODOM BLAST-DOMO
12	517	S485 S4 T11 S128 T133 S155 S156 S171 S172 S278 T288 S485 S3 T57 T199 T204 S278 T455 S462 T480	N286	Ubiquitin carboxyl-terminal hydrolases family 2: K61-P256, F436-V481, S470-S491	g2459395 Ubiquitin protease	Motifs BLAST-GenBank HMMER-PFAM BLIMPS-BLOCKS BLAST-PRODOM
13	346	T237 S12 T64 T72 T124 T236 T261 S319 S150 T194 S226 T251 S319		Ubiquitin-activating enzyme signature: S297-344, 11-163, 19-189, 7-174, 9-192, R35-G170 Membrane protein: 11-249	g3647283 Ubiquitin activating enzyme	Motifs BLAST-GenBank HMMER BLAST-PRODOM BLAST-DOMO
14	151	T24 T47 S118 S61 Y131		Ubiquitin-conjugating enzymes: M1-D148 Active site: F58-M115	g4090259 Ubiquitin-conjugating enzyme E2	Motifs BLAST-GenBank HMMER-PFAM ProfileScan BLAST-PRODOM BLAST-DOMO
15	362	S199 S208 S212 S270 S281 T317 S327 S52 S122 T149	N120 N162 N175 N239	Signal peptide: M1-S26 Zinc carboxypeptidase: Y38-E320 Zinc binding region: E202-L258	g6013463 Carboxypeptidase homolog	Motifs BLAST-GenBank SPScan HMMER-PFAM ProfileScan BLAST-PRODOM BLAST-DOMO

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences	Homologous Sequence	Analytical Methods and Databases
16	123	S2 S9 T37 T46 T60 S112 T53 S112	N104	Kunitz type protease inhibitor active site region: C70-C120	g512802 Kunitz type protease inhibitor	Motifs BLAST-GenBank HMMER-PFAM ProfileScan BLAST-DOMO
17	983	S87 S461 S531 T761 T123 T143 S191 S445 S634 S660 T789 T820 S879 S886 S888 T890 T17 S158 T280 T398 T549 S598 S601 S687 Y268 Y688	N278 N427 N625 N884 N922	Ubiquitin carboxyl-terminal hydrolases family 2: G90-w107, Y336-I374	g1429371 Ubiquitin-specific protease	Motifs BLAST-GenBank HMMER-PFAM BLAST-DOMO
18	227	S49 T101 T131 T157 S166 S49 S144 S194 T199		Ubiquitin signature: K159-H179, A180-D200 (P value = 0.00032)	g9372 Ubiquitin (P value = $1.7e^{-08}$)	Motifs BLAST-GenBank BLIMPS-PRINTS
19	403	T47 S146 T261 T352 T381 S4 T119 S234 S291 S313	N117 N145 N232 N260 N289 N317	Ubiquitin carboxyl-terminal hydrolases family 2: G221-L238	g4731026 Nod1 activator of caspase-9 and NFKB	Motifs BLAST-GenBank HMMER-PFAM
20	372	T87 S291 S22 S197 T208 S343 T169 S185 S223 S260 T266	N188 N335	Ubiquitin carboxyl-terminal hydrolases family 2: A166-Q348 Active site: Y302-C320	g4469352 Ubiquitin specific protease UBP43	Motifs BLAST-GenBank HMMER-PFAM BLAST-DOMO
21	94	T9	N50	Signal peptidase: V41-R55	g3687497 Putative mitochondrial inner membrane protease subunit	Motifs BLAST-GenBank

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences	Homologous Sequence	Analytical Methods and Databases
22	248	S77 S135 S156 S183 S205 T3 S71 S72 T139	N47 N158	Alpha-2-macroglobulin family: T3-Y198 Complement precursor: E4-S206	g2073373 Alpha-2-macroglobulin protease inhibitor	Motifs BLAST-GenBank HMMER-PFAM BLAST-PRODOM
23	520	S166 S272 T301 S326 S379 S455 S56 T82 S136 S227 S498	N164 N355	Signal peptide: M1-R27 Peptidase M10: F39-S225 Matrixin domain: F128-G288 Neutral zinc metallopeptidase zinc-binding region: V237-L246 Hemopexin domain: I341-K400	g1731986 MMP-19 matrix metalloproteinase	Motifs BLAST-GenBank SIGPEPT SPScan HMMER-PFAM ProfileScan BLAST-PRODOM BLAST-DOMO
24	422	T188 S156 S306 T386 S130 T176 T226 T295 S357 S365	N94 N106 N169 N350	Signal peptide: M1-G26 Transmembrane domain: F398-N418 Serpins (serine protease inhibitors): P43-V420 Protease "bait" region: A371-G422	g425146 Kallistatin	Motifs BLAST-GenBank SIGPEPT SPScan HMMER HMMER-PFAM BLIMPS-BLOCKS ProfileScan BLAST-PRODOM BLAST-DOMO
25	114	S74 S16		Eukaryotic thiol (cysteine) protease active site: R71-S114		Motifs BLAST-GenBank ProfileScan

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences	Homologous Sequence	Analytical Methods and Databases
26	742	T167 S186 S308 S337 S343 T360 S439 S578 S92 S172 S239 T256 T278 S329 T414 S504 S633 T656 T708 Y28 Y107 Y356		Zinc carboxypeptidases, zinc-binding regions signatures: H32-W42		Motifs BLAST-GenBank
27	734	T83 S128 S151 S223 S233 T523 S574 T616 T665 T688 T34 S122 S203 S340 T546 S547 T703	N57 N210 N220 N318 N428 N472	Signal peptide: M1-G20 Zinc carboxypeptidases: H299-Y412, W421-Y678 Enkephalin convertase: P458-V687 Zinc binding region: E478-F529	g4322263 Metallocarboxy-peptidase CPX-1	Motifs BLAST-GenBank SIGPEPT SPScan HMMER-PFAM ProfileScan BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO

Table 3

Nucleotide SEQ ID NO:	Selected Fragment	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
28	164-208	Gastrointestinal (1.000)	Inflammation (0.500)	PBLUESCRIPT
29	57-101	Reproductive (0.274) Nervous (0.202) Cardiovascular (0.119) Gastrointestinal (0.119)	Cancer (0.524) Inflammation (0.273) Cell Proliferation(0.190)	PBLUESCRIPT
30	111-155	Nervous (0.222) Reproductive (0.194) Gastrointestinal (0.139)	Cancer (0.403) Inflammation (0.361)	PSPORT1
31	921-965	Nervous (0.300) Reproductive (0.200) Cardiovascular (0.100) Dermatologic (0.100) Developmental (0.100) Gastrointestinal (0.100) Hematopoietic/Immune (0.100)	Cancer (0.400) Cell Proliferation(0.300) Neurological (0.100)	pINCY
32	809-853	Hematopoietic/Immune (0.194) Reproductive (0.239) Gastrointestinal (0.164)	Cancer (0.403) Inflammation (0.269) Cell Proliferation(0.134)	pINCY
33	273-317	Reproductive (0.500) Cardiovascular (0.125) Dermatologic (0.125) Gastrointestinal (0.125) Hematopoietic/Immune (0.125)	Cancer (0.625) Cell Proliferation(0.125) Inflammation (0.125)	pINCY
34	55-99	Nervous (0.185) Cardiovascular (0.111) Gastrointestinal (0.111)	Cancer (0.352) Inflammation (0.204) Cell Proliferation(0.204)	pINCY
35	218-262	Gastrointestinal (0.313) Hematopoietic/Immune (0.250) Reproductive (0.188)	Cancer (0.630) Cell Proliferation(0.250)	PSPORT1
36	325-369	Developmental (0.500) Gastrointestinal (0.500)	Cancer (0.500) Cell Proliferation(0.500)	PSPORT1
37	99-143	Nervous (0.198) Reproductive (0.165) Cardiovascular (0.154)	Cancer (0.374) Inflammation (0.374) Cell Proliferation(0.154)	pINCY

Table 3 (cont.)

Nucleotide SEQ ID NO:	Selected Fragment	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
38	1-46	Reproductive (0.278) Gastrointestinal (0.208) Cardiovascular (0.125)	Cancer (0.347) Inflammation (0.306) Cell Proliferation(0.153)	pINCY
39	109-153	Gastrointestinal (0.280) Hematopoietic/Immune (0.200) Musculoskeletal (0.120)	Inflammation (0.440) Cancer (0.280) Cell Proliferation(0.160)	pINCY
40	489-533	Nervous (0.209) Reproductive (0.203) Gastrointestinal (0.135)	Cancer (0.473) Cell Proliferation(0.243) Inflammation (0.264	pINCY
41	589-633	Reproductive (0.229) Cardiovascular (0.200) Gastrointestinal (0.171) Nervous (0.171)	Cancer (0.314) Cell Proliferation(0.314) Inflammation/Trauma (0.372)	PBLUESCRIPT
42	649-693	Nervous (0.250) Reproductive (0.214) Cardiovascular (0.143)	Cancer (0.500) Inflammation/Trauma (0.321) Cell Proliferation(0.179)	pINCY
43	164-208	Nervous (0.333) Gastrointestinal (0.333) Reproductive (0.333)	Cancer (0.444) Inflammation/Trauma (0.444) Neurological (0.111)	pINCY
44	271-208	Reproductive (0.226) Developmental (0.151) Nervous (0.151)	Cancer (0.377) Inflammation/Trauma (0.358) Cell Proliferation(0.321)	pINCY
45	784-828	Reproductive (0.257) Hematopoietic/Immune (0.171) Nervous (0.171)	Cancer (0.486) Inflammation/Trauma (0.486) Cell Proliferation(0.143)	PBLUESCRIPT
46	219-263	Reproductive (0.444) Gastrointestinal (0.222) Nervous (0.222)	Inflammation/Trauma (0.666) Cancer (0.222)	PSPORT1
47	597-641	Reproductive (0.364) Cardiovascular (0.212) Nervous (0.152)	Cancer (0.545) Cell Proliferation(0.242) Inflammation/Trauma (0.273)	PSPORT1

Table 3 (cont.)

Nucleotide SEQ ID NO:	Selected Fragment	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
48	271-315	Gastrointestinal (0.278) Reproductive (0.278) Cardiovascular (0.111) Hematopoietic/Immune (0.111) Nervous (0.111)	Cancer (0.444) Inflammation/Trauma (0.555) Cell Proliferation(0.167)	pINCY
49	217-261	Hematopoietic/Immune (0.364) Reproductive (0.273)	Cell Proliferation(0.364) Inflammation/Trauma (0.364) Cancer (0.182)	pINCY
50	164-208	Reproductive (0.333) Nervous (0.222) Gastrointestinal (0.167)	Cancer (0.611) Inflammation/Trauma (0.223)	pINCY
51	388-432	Gastrointestinal (0.833) Reproductive (0.166)	Cancer (0.666) Cell Proliferation(0.166)	pINCY
52	218-262	Nervous (0.750) Hematopoietic/Immune (0.250)	Inflammation/Trauma (0.500) Neurological (0.250)	pINCY
53	325-369	Reproductive (0.289) Nervous (0.253) Gastrointestinal (0.120)	Cancer (0.410) Inflammation/Trauma (0.386) Cell Proliferation(0.145)	pINCY
54	165-209	Reproductive (0.352) Urologic (0.185) Developmental (0.130)	Cancer (0.630) Cell Proliferation(0.167) Inflammation/Trauma (0.204)	pINCY

Table 4

Polynucleotide SEQ ID NO:	Library	Library Comment
28	LIVRNOT01	Library was constructed at Stratagene, using RNA isolated from the liver tissue of a 49-year-old male.
29	TESTNOT01	Library was constructed using RNA isolated from the testicular tissue of a 37-year-old Caucasian male, who died from liver disease. Patient history included cirrhosis, jaundice, and liver failure.
30	SYNORAT05	Library was constructed using RNA isolated from the knee synovial tissue of a 62-year-old female with rheumatoid arthritis.
31	BRSTNOT07	Library was constructed using RNA isolated from diseased breast tissue removed from a 43-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated mildly proliferative fibrocystic changes with epithelial hyperplasia, papillomatosis, and duct ectasia. Pathology for the associated tumor tissue indicated invasive grade 4, nuclear grade 3 mammary adenocarcinoma with extensive comedo necrosis. Family history included epilepsy, cardiovascular disease, and type II diabetes.
32	LEUKNOT02	Library was constructed using RNA isolated from white blood cells of a 45-year-old female with blood type O+. The donor tested positive for cytomegalovirus (CMV).
33	OVARNOT07	Library was constructed using RNA isolated from left ovarian tissue removed from a 28-year-old Caucasian female during a vaginal hysterectomy and removal of the fallopian tubes and ovaries. The tissue was associated with multiple follicular cysts, endometrium in a weakly proliferative phase, and chronic cervicitis of the cervix with squamous metaplasia. Family history included benign hypertension, hyperlipidemia, and atherosclerotic coronary artery disease.
34	CORPNOT02	Library was constructed using RNA isolated from diseased corpus callosum tissue removed from the brain of a 74-year-old Caucasian male who died from Alzheimer's disease.
35	BRAINON01	Library was constructed and normalized from 4.88 million independent clones from a brain tissue library. RNA was made from brain tissue removed from a 26-year-old Caucasian male during cranioplasty and excision of a cerebral meningeal lesion. Pathology for the associated tumor tissue indicated a grade 4 oligoastrocytoma in the right fronto-parietal part of the brain.
36	COLNNOT11	Library was constructed using RNA isolated from colon tissue removed from a 60-year-old Caucasian male during a left hemicolectomy.

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
37	BRAITUT21	Library was constructed using RNA isolated from brain tumor tissue removed from the midline frontal lobe of a 61-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated subfrontal meningothelial meningioma with no atypia. One ethmoid and mucosal tissue sample indicated meningioma. Family history included cerebrovascular disease, senile dementia, hyperlipidemia, benign hypertension, atherosclerotic coronary artery disease, congestive heart failure, and breast cancer.
38	EPIGNOT01	Library was constructed using RNA isolated from epiglottic tissue removed from a 71-year-old male during laryngectomy with right parathyroid biopsy. Pathology for the associated tumor tissue indicated recurrent grade 1 papillary thyroid carcinoma.
39	SPLNNOT17	Library was constructed using polyA RNA isolated from the spleen tissue of a 2-year-old Hispanic male who died from cerebral anoxia.
40	FIBAUNT02	Library was constructed using RNA isolated from untreated aortic adventitial fibroblasts removed from a 65-year-old Caucasian female.
41	BRSTNOT01	Library was constructed using RNA isolated from the breast tissue of a 56-year-old Caucasian female who died in a motor vehicle accident.
42	MENITUT03	Library was constructed using RNA isolated from brain meningioma tissue removed from a 35-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated a benign neoplasm in the right cerebellopontine angle of the brain. Patient history included hypothyroidism. Family history included myocardial infarction and breast cancer .
43	BRAINOT10	Library was constructed using RNA isolated from diseased cerebellum tissue removed from the brain of a 74-year-old Caucasian male, who died from Alzheimer's disease.
44	CONNTUT01	Library was constructed using RNA isolated from a soft tissue tumor removed from the clival area of the skull of a 30-year-old Caucasian female. Pathology indicated chondroid chordoma with neoplastic cells reactive for keratin.
45	UCMCL5T01	Library was constructed using RNA isolated from mononuclear cells obtained from the umbilical cord blood of 12 individuals. The cells were cultured for 12 days with IL-5 before RNA was obtained from the pooled lysates.
46	OVARNOT02	Library was constructed using RNA isolated from ovarian tissue removed from a 59-year-old Caucasian female who died of a myocardial infarction. Patient history included cardiomyopathy, coronary artery disease, previous myocardial infarctions, hypercholesterolemia, hypotension, and arthritis.

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
47	THP1AZS08	This subtracted THP-1 promonocyte cell line library was constructed using 5.76 million clones from a 5-aza-2'-deoxycytidine (AZ) treated THP-1 cell library. Starting RNA was made from THP-1 promonocyte cells treated for three days with 0.8 micromolar AZ. The hybridization probe for subtraction was derived from a similarly constructed library, made from RNA isolated from untreated THP-1 cells. 5.76 million clones from the AZ-treated THP-1 cell library were then subjected to two rounds of subtractive hybridization with 5 million clones from the untreated THP-1 cell library. Subtractive hybridization conditions were based on the methodologies of Swaroop et al. (1991) Nucleic Acids Res. 19:1954; and Bonaldo et al. (1996) Genome Research 6:791. THP-1 (ATCC TIB 202) is a human promonocyte line derived from peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia.
48	BRSTNOT19	Library was constructed using RNA isolated from breast tissue removed from a 67-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated residual invasive lobular carcinoma. The focus of residual invasive carcinoma was positive for both estrogen and progesterone. Patient history included depressive disorder and benign large bowel neoplasm. Family history included cerebrovascular disease, benign hypertension, congestive heart failure, and lung cancer.
49	ENDPNOT02	Library was constructed using RNA isolated from pulmonary artery endothelial cells removed from a 10-year-old Caucasian male. The cells were treated with TNF alpha and IL-1 beta 10ng/ml each for 20 hours.
50	UTRSNOT05	Library was constructed using RNA isolated from the uterine tissue of a 45-year-old Caucasian female during a total abdominal hysterectomy and total colectomy. Pathology for the associated tumor tissue indicated multiple leiomyomas of the myometrium and a grade 2 colonic adenocarcinoma of the cecum. Patient history included multiple sclerosis and mitral valve disorder. Family history included type I diabetes, cerebrovascular disease, atherosclerotic coronary artery disease, malignant skin neoplasm, hypertension, and malignant neoplasm of the colon.

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
51	LIVRTUT11	Library was constructed using 1.1 micrograms of polyA RNA isolated from a treated C3A hepatocyte cell line which is a derivative of Hep G2, a cell line derived from a hepatoblastoma removed from a 15-year-old Caucasian male. The cells were treated with phenobarbital (PB), 1mM for 48 hours. cDNA synthesis was initiated using a NotI-anchored oligo(dT) primer. Double-stranded cDNA was blunted, ligated to EcoRI adaptors, digested with NotI, size-selected, and cloned into the NotI and EcoRI sites of the pINCY vector (Incyte).
52	BRAFNOT02	The library was constructed using RNA isolated from superior frontal cortex tissue removed from a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly, and an enlarged spleen and liver.
53	BRAFDIT02	The library was constructed using RNA isolated from diseased right frontal lobe tissue removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident. Patient history included Huntington's disease and emphysema.
54	FIBAUNT01	Library was constructed using RNA isolated from untreated aortic adventitial fibroblasts obtained from a 48-year-old Caucasian male.

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	PE Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	PE Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	PE Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and, if applicable, Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits for PFAM hits, depending on individual protein families

Table 5 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

a) an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, and SEQ ID NO:27,

b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, and SEQ ID NO:27,

c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, and SEQ ID NO:27, and

d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, and SEQ ID NO:27.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, and SEQ ID NO:27.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, and SEQ ID NO:54.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method for producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, and SEQ ID NO:54,

b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a

polynucleotide sequence selected from the group consisting of SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, and SEQ ID NO:54,

c) a polynucleotide sequence complementary to a),

d) a polynucleotide sequence complementary to b), and

e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4,

SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, and SEQ ID NO:27.

18. A method for treating a disease or condition associated with decreased expression of functional PPIM, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional PPIM, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional PPIM, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

a) combining the polypeptide of claim 1 with at least one test compound under suitable

conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, and

b) detecting altered expression of the target polynucleotide.

28. A method for assessing toxicity of a test compound, said method comprising:

a) treating a biological sample containing nucleic acids with the test compound;

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;

c) quantifying the amount of hybridization complex; and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

SEQUENCE LISTING

<110> INCYTE GENOMICS, INC.
VUE, Henry
TANG, Y. Tom
BANDMAN, Olga
LAL, Preeti
BAUGHN, Mariah R.
AZIMZAI, Yalda
LU, Dyung Aina M.
YANG, Junming

<120> PROTEASES AND PROTEASE INHIBITORS

<130> PF-0727 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/147,986; 60/160,807
<151> 1999-08-09; 1999-10-21

<160> 54
<170> PERL Program

<210> 1
<211> 444
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 088718CD1

<400> 1
Met Lys Val Val Pro Ser Leu Leu Leu Ser Val Leu Leu Ala Gln
1 5 10 15
Val Trp Leu Val Pro Gly Leu Ala Pro Ser Pro Gln Ser Pro Glu
20 25 30
Thr Pro Ala Pro Gln Asn Gln Thr Ser Arg Val Val Gln Ala Pro
35 40 45
Arg Glu Glu Glu Glu Asp Glu Gln Glu Ala Ser Glu Glu Lys Ala
50 55 60
Gly Glu Glu Glu Lys Ala Trp Leu Met Ala Ser Arg Gln Gln Leu
65 70 75
Ala Lys Glu Thr Ser Asn Phe Gly Phe Ser Leu Leu Arg Lys Ile
80 85 90
Ser Met Arg His Asp Gly Asn Met Val Phe Ser Pro Phe Gly Met
95 100 105
Ser Leu Ala Met Thr Gly Leu Met Leu Gly Ala Thr Gly Pro Thr
110 115 120
Glu Thr Gln Ile Lys Arg Gly Leu His Leu Gln Ala Leu Lys Pro
125 130 135
Thr Lys Pro Gly Leu Leu Pro Ser Leu Phe Lys Gly Leu Arg Glu
140 145 150
Thr Leu Ser Arg Asn Leu Glu Leu Gly Leu Ser Gln Gly Ser Phe
155 160 165
Ala Phe Ile His Lys Asp Phe Asp Val Lys Glu Thr Phe Phe Asn
170 175 180
Leu Ser Lys Arg Tyr Phe Asp Thr Glu Cys Val Pro Met Asn Phe
185 190 195
Arg Asn Ala Ser Gln Ala Lys Arg Leu Met Asn His Tyr Ile Asn
200 205 210
Lys Glu Thr Arg Gly Lys Ile Pro Lys Leu Phe Asp Glu Ile Asn
215 220 225
Pro Glu Thr Lys Leu Ile Leu Val Asp Tyr Ile Leu Phe Lys Gly

```
                230                 235                 240
Lys Trp Leu Thr Pro Phe Asp Pro Val Phe Thr Glu Val Asp Thr
                245                 250                 255
Phe His Leu Asp Lys Tyr Lys Thr Ile Lys Val Pro Met Met Tyr
                260                 265                 270
Gly Ala Gly Lys Phe Ala Ser Thr Phe Asp Lys Asn Phe Arg Cys
                275                 280                 285
His Val Leu Lys Leu Pro Tyr Gln Gly Asn Ala Thr Met Leu Val
                290                 295                 300
Val Leu Met Glu Lys Met Gly Asp His Leu Ala Leu Glu Asp Tyr
                305                 310                 315
Leu Thr Thr Asp Leu Val Glu Thr Trp Leu Arg Asn Met Lys Thr
                320                 325                 330
Arg Asn Met Glu Val Phe Phe Pro Lys Phe Lys Leu Asp Gln Lys
                335                 340                 345
Tyr Glu Met His Glu Leu Leu Arg Gln Met Gly Ile Arg Arg Ile
                350                 355                 360
Phe Ser Pro Phe Ala Asp Leu Ser Glu Leu Ser Ala Thr Gly Arg
                365                 370                 375
Asn Leu Gln Val Ser Arg Val Leu Gln Arg Thr Val Ile Glu Val
                380                 385                 390
Asp Glu Arg Gly Thr Glu Ala Val Ala Gly Ile Leu Ser Glu Ile
                395                 400                 405
Thr Ala Tyr Ser Met Pro Pro Val Ile Lys Val Asp Arg Pro Phe
                410                 415                 420
His Phe Met Ile Tyr Glu Glu Thr Ser Gly Met Leu Leu Phe Leu
                425                 430                 435
Gly Arg Val Val Asn Pro Thr Leu Leu
                440
<210> 2
<211> 565
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 114551CD1

<400> 2
Met Ser Gly Arg Ser Lys Arg Glu Ser Arg Gly Ser Thr Arg Gly
  1               5                  10                  15
Lys Arg Glu Ser Glu Ser Arg Gly Ser Ser Gly Arg Val Lys Arg
             20                  25                  30
Glu Arg Asp Arg Glu Arg Glu Pro Glu Ala Ala Ser Ser Arg Gly
             35                  40                  45
Ser Pro Val Arg Val Lys Arg Glu Phe Glu Pro Ala Ser Ala Arg
             50                  55                  60
Glu Ala Pro Ala Ser Val Val Pro Phe Val Arg Val Lys Arg Glu
             65                  70                  75
Arg Glu Val Asp Glu Asp Ser Glu Pro Glu Arg Glu Val Arg Ala
             80                  85                  90
Lys Asn Gly Arg Val Asp Ser Glu Asp Arg Arg Ser Arg His Cys
             95                 100                 105
Pro Tyr Leu Asp Thr Ile Asn Arg Ser Val Leu Asp Phe Asp Phe
            110                 115                 120
Glu Lys Leu Cys Ser Ile Ser Leu Ser His Ile Asn Ala Tyr Ala
            125                 130                 135
Cys Leu Val Cys Gly Lys Tyr Phe Gln Gly Arg Gly Leu Lys Ser
            140                 145                 150
His Ala Tyr Ile His Ser Val Gln Phe Ser His His Val Phe Leu
            155                 160                 165
Asn Leu His Thr Leu Lys Phe Tyr Cys Leu Pro Asp Asn Tyr Glu
            170                 175                 180
Ile Ile Asp Ser Ser Leu Glu Asp Ile Thr Tyr Val Leu Lys Pro
            185                 190                 195
Thr Phe Thr Lys Gln Gln Ile Ala Asn Leu Asp Lys Gln Ala Lys
            200                 205                 210
```

Leu Ser Arg Ala Tyr Asp Gly Thr Thr Tyr Leu Pro Gly Ile Val
215 220 225
Gly Leu Asn Asn Ile Lys Ala Asn Asp Tyr Ala Asn Ala Val Leu
230 235 240
Gln Ala Leu Ser Asn Val Pro Pro Leu Arg Asn Tyr Phe Leu Glu
245 250 255
Glu Asp Asn Tyr Lys Asn Ile Lys Arg Pro Pro Gly Asp Ile Met
260 265 270
Phe Leu Leu Val Gln Arg Phe Gly Glu Leu Met Arg Lys Leu Trp
275 280 285
Asn Pro Arg Asn Phe Lys Ala His Val Ser Pro His Glu Met Leu
290 295 300
Gln Ala Val Val Leu Cys Ser Lys Lys Thr Phe Gln Ile Thr Lys
305 310 315
Gln Gly Asp Gly Val Asp Phe Leu Ser Trp Phe Leu Asn Ala Leu
320 325 330
His Ser Ala Leu Gly Gly Thr Lys Lys Lys Lys Lys Thr Ile Val
335 340 345
Thr Asp Val Phe Gln Gly Ser Met Arg Ile Phe Thr Lys Lys Leu
350 355 360
Pro His Pro Asp Leu Pro Ala Glu Glu Lys Glu Gln Leu Leu His
365 370 375
Asn Asp Glu Tyr Gln Glu Thr Met Val Glu Ser Thr Phe Met Tyr
380 385 390
Leu Thr Leu Asp Leu Pro Thr Ala Pro Leu Tyr Lys Asp Glu Lys
395 400 405
Glu Gln Leu Ile Ile Pro Gln Val Pro Leu Phe Asn Ile Leu Ala
410 415 420
Lys Phe Asn Gly Ile Thr Glu Lys Glu Tyr Lys Thr Tyr Lys Glu
425 430 435
Asn Phe Leu Lys Arg Phe Gln Leu Thr Lys Leu Pro Pro Tyr Leu
440 445 450
Ile Phe Cys Ile Lys Arg Phe Thr Lys Asn Asn Phe Phe Val Glu
455 460 465
Lys Asn Pro Thr Ile Val Asn Phe Pro Ile Thr Asn Val Asp Leu
470 475 480
Arg Glu Tyr Leu Ser Glu Glu Val Gln Ala Val His Lys Asn Thr
485 490 495
Thr Tyr Asp Leu Ile Ala Asn Ile Val His Asp Gly Lys Pro Ser
500 505 510
Glu Gly Ser Tyr Arg Ile His Val Leu His His Gly Thr Gly Lys
515 520 525
Trp Tyr Glu Leu Gln Asp Leu Gln Val Thr Asp Ile Leu Pro Gln
530 535 540
Met Ile Thr Leu Ser Glu Ala Tyr Ile Gln Ile Trp Lys Arg Arg
545 550 555
Asp Asn Asp Glu Thr Asn Gln Gln Gly Ala
560 565

<210> 3
<211> 589
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1261376CD1

<400> 3
Met Ala Glu Ser Gly Glu Ser Gly Gly Pro Pro Gly Ser Gln Asp
1 5 10 15
Ser Ala Ala Gly Ala Glu Gly Ala Gly Ala Pro Ala Ala Ala Ala
20 25 30
Ser Ala Asp Ala Lys Ile Met Lys Val Thr Val Lys Thr Pro Lys
35 40 45
Glu Lys Glu Glu Phe Ala Val Pro Glu Asn Ser Ser Val Gln Gln
50 55 60
Phe Lys Glu Glu Ile Ser Lys Arg Phe Lys Ser His Thr Asp Gln

```
                    65                  70                  75
Leu Val Leu Ile Phe Ala Gly Lys Ile Leu Lys Asp Gln Asp Thr
                    80                  85                  90
Leu Ser Gln His Gly Ile His Asp Gly Leu Thr Val His Leu Val
                    95                 100                 105
Ile Lys Thr Gln Asn Arg Pro Gln Asp His Ser Ala Gln Gln Thr
                   110                 115                 120
Asn Thr Ala Gly Ser Asn Val Thr Thr Ser Ser Thr Pro Asn Ser
                   125                 130                 135
Asn Ser Thr Ser Gly Ser Ala Thr Ser Asn Pro Phe Gly Leu Gly
                   140                 145                 150
Gly Leu Gly Gly Leu Ala Gly Leu Ser Ser Leu Gly Leu Asn Thr
                   155                 160                 165
Thr Asn Phe Ser Glu Leu Gln Ser Gln Met Gln Arg Gln Leu Leu
                   170                 175                 180
Ser Asn Pro Glu Met Met Val Gln Ile Met Glu Asn Pro Phe Val
                   185                 190                 195
Gln Ser Met Leu Ser Asn Pro Asp Leu Met Arg Gln Leu Ile Met
                   200                 205                 210
Ala Asn Pro Gln Met Gln Gln Leu Ile Gln Arg Asn Pro Glu Ile
                   215                 220                 225
Ser His Met Leu Asn Asn Pro Asp Ile Met Arg Gln Thr Leu Glu
                   230                 235                 240
Leu Ala Arg Asn Pro Ala Met Met Gln Glu Met Met Arg Asn Gln
                   245                 250                 255
Asp Arg Ala Leu Ser Asn Leu Glu Ser Ile Pro Gly Gly Tyr Asn
                   260                 265                 270
Ala Leu Arg Arg Met Tyr Thr Asp Ile Gln Glu Pro Met Leu Ser
                   275                 280                 285
Ala Ala Gln Glu Gln Phe Gly Gly Asn Pro Phe Ala Ser Leu Val
                   290                 295                 300
Ser Asn Thr Ser Ser Gly Glu Gly Ser Gln Pro Ser Arg Thr Glu
                   305                 310                 315
Asn Arg Asp Pro Leu Pro Asn Pro Trp Ala Pro Gln Thr Ser Gln
                   320                 325                 330
Ser Ser Ser Ala Ser Ser Gly Thr Ala Ser Thr Val Gly Gly Thr
                   335                 340                 345
Thr Gly Ser Thr Ala Ser Gly Thr Ser Gly Gln Ser Thr Thr Ala
                   350                 355                 360
Pro Asn Leu Val Pro Gly Val Gly Ala Ser Met Phe Asn Thr Pro
                   365                 370                 375
Gly Met Gln Ser Leu Leu Gln Gln Ile Thr Glu Asn Pro Gln Leu
                   380                 385                 390
Met Gln Asn Met Leu Ser Ala Pro Tyr Met Arg Ser Met Met Gln
                   395                 400                 405
Ser Leu Ser Gln Asn Pro Asp Leu Ala Ala Gln Met Met Leu Asn
                   410                 415                 420
Asn Pro Leu Phe Ala Gly Asn Pro Gln Leu Gln Glu Gln Met Arg
                   425                 430                 435
Gln Gln Leu Pro Thr Phe Leu Gln Gln Met Gln Asn Pro Asp Thr
                   440                 445                 450
Leu Ser Ala Met Ser Asn Pro Arg Ala Met Gln Ala Leu Leu Gln
                   455                 460                 465
Ile Gln Gln Gly Leu Gln Thr Leu Ala Thr Glu Ala Pro Gly Leu
                   470                 475                 480
Ile Pro Gly Phe Thr Pro Gly Leu Gly Ala Leu Gly Ser Thr Gly
                   485                 490                 495
Gly Ser Ser Gly Thr Asn Gly Ser Asn Ala Thr Pro Ser Glu Asn
                   500                 505                 510
Thr Ser Pro Thr Ala Gly Thr Thr Glu Pro Gly His Gln Gln Phe
                   515                 520                 525
Ile Gln Gln Met Leu Gln Ala Leu Ala Gly Val Asn Pro Gln Leu
                   530                 535                 540
Gln Asn Pro Glu Val Arg Phe Gln Gln Gln Leu Glu Gln Leu Ser
                   545                 550                 555
Ala Met Gly Phe Leu Asn Arg Glu Ala Asn Leu Gln Ala Leu Ile
                   560                 565                 570
```

```
Ala Thr Gly Gly Asp Ile Asn Ala Ala Ile Glu Arg Leu Leu Gly
                575                 580                 585
Ser Gln Pro Ser

<210> 4
<211> 775
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1299481CD1

<400> 4
Met Thr Ile Val Asp Lys Ala Ser Glu Ser Ser Asp Pro Ser Ala
  1               5                  10                  15
Tyr Gln Asn Gln Pro Gly Ser Ser Glu Ala Val Ser Pro Gly Asp
             20                  25                  30
Met Asp Ala Gly Ser Ala Ser Trp Gly Ala Val Ser Ser Leu Asn
         35                  40                  45
Asp Val Ser Asn His Thr Leu Ser Leu Gly Pro Val Pro Gly Ala
     50                  55                  60
Val Val Tyr Ser Ser Ser Ser Val Pro Asp Lys Ser Lys Pro Ser
 65                  70                  75
Pro Gln Lys Asp Gln Ala Leu Gly Asp Gly Ile Ala Pro Pro Gln
             80                  85                  90
Lys Val Leu Phe Pro Ser Glu Lys Ile Cys Leu Lys Trp Gln Gln
         95                 100                 105
Thr His Arg Val Gly Ala Gly Leu Gln Asn Leu Gly Asn Thr Cys
    110                 115                 120
Phe Ala Asn Ala Ala Leu Gln Cys Leu Thr Tyr Thr Pro Pro Leu
    125                 130                 135
Ala Asn Tyr Met Leu Ser His Glu His Ser Lys Thr Cys His Ala
    140                 145                 150
Glu Gly Phe Cys Met Met Cys Thr Met Gln Ala His Ile Thr Gln
    155                 160                 165
Ala Leu Ser Asn Pro Gly Asp Val Ile Lys Pro Met Phe Val Ile
    170                 175                 180
Asn Glu Met Arg Arg Ile Ala Arg His Leu Arg Phe Gly Asn Gln
    185                 190                 195
Glu Asp Ala His Glu Phe Leu Gln Tyr Thr Val Asp Ala Met Gln
    200                 205                 210
Lys Ala Cys Leu Asn Gly Ser Asn Lys Leu Asp Arg His Thr Gln
    215                 220                 225
Ala Thr Thr Leu Val Cys Gln Ile Phe Gly Gly Tyr Leu Arg Ser
    230                 235                 240
Arg Val Lys Cys Leu Asn Cys Lys Gly Val Ser Asp Thr Phe Asp
    245                 250                 255
Pro Tyr Leu Asp Ile Thr Leu Glu Ile Lys Ala Ala Gln Ser Val
    260                 265                 270
Asn Lys Ala Leu Glu Gln Phe Val Lys Pro Glu Gln Leu Asp Gly
    275                 280                 285
Glu Asn Ser Tyr Lys Cys Ser Lys Cys Lys Lys Met Val Pro Ala
    290                 295                 300
Ser Lys Arg Phe Thr Ile His Arg Ser Ser Asn Val Leu Thr Leu
    305                 310                 315
Ser Leu Lys Arg Phe Ala Asn Phe Thr Gly Gly Lys Ile Ala Lys
    320                 325                 330
Asp Val Lys Tyr Pro Glu Tyr Leu Asp Ile Arg Pro Tyr Met Ser
    335                 340                 345
Gln Pro Asn Gly Glu Pro Ile Val Tyr Val Leu Tyr Ala Val Leu
    350                 355                 360
Val His Thr Gly Phe Asn Cys His Ala Gly His Tyr Phe Cys Tyr
    365                 370                 375
Ile Lys Ala Ser Asn Gly Leu Trp Tyr Gln Met Asn Asp Ser Ile
    380                 385                 390
Val Ser Thr Ser Asp Ile Arg Ser Val Leu Ser Gln Gln Ala Tyr
```

```
                395                 400                 405
Val Leu Phe Tyr Ile Arg Ser His Asp Val Lys Asn Gly Gly Glu
                410                 415                 420
Leu Thr His Pro Thr His Ser Pro Gly Gln Ser Ser Pro Arg Pro
                425                 430                 435
Val Ile Ser Gln Arg Val Val Thr Asn Lys Gln Ala Ala Pro Gly
                440                 445                 450
Phe Ile Gly Pro Gln Leu Pro Ser His Met Ile Lys Asn Pro Pro
                455                 460                 465
His Leu Asn Gly Thr Gly Pro Leu Lys Asp Thr Pro Ser Ser Ser
                470                 475                 480
Met Ser Ser Pro Asn Gly Asn Ser Ser Val Asn Arg Ala Ser Pro
                485                 490                 495
Val Asn Ala Ser Ala Ser Val Gln Asn Trp Ser Val Asn Arg Ser
                500                 505                 510
Ser Val Ile Pro Glu His Pro Lys Lys Gln Lys Ile Thr Ile Ser
                515                 520                 525
Ile His Asn Lys Leu Pro Val Arg Gln Cys Gln Ser Gln Pro Asn
                530                 535                 540
Leu His Ser Asn Ser Leu Glu Asn Pro Thr Lys Pro Val Pro Ser
                545                 550                 555
Ser Thr Ile Thr Asn Ser Ala Val Gln Ser Thr Ser Asn Ala Ser
                560                 565                 570
Thr Met Ser Val Ser Ser Lys Val Thr Lys Pro Ile Pro Arg Ser
                575                 580                 585
Glu Ser Cys Ser Gln Pro Val Met Asn Gly Lys Ser Lys Leu Asn
                590                 595                 600
Ser Ser Val Leu Val Pro Tyr Gly Ala Glu Ser Ser Glu Asp Ser
                605                 610                 615
Asp Glu Glu Ser Lys Gly Leu Gly Lys Glu Asn Gly Ile Gly Thr
                620                 625                 630
Ile Val Ser Ser His Ser Pro Gly Gln Asp Ala Glu Asp Glu Glu
                635                 640                 645
Ala Thr Pro His Glu Leu Gln Glu Pro Met Thr Leu Asn Gly Ala
                650                 655                 660
Asn Ser Ala Asp Ser Asp Ser Asp Pro Lys Glu Asn Gly Leu Ala
                665                 670                 675
Pro Asp Gly Ala Ser Cys Gln Gly Gln Pro Ala Leu His Ser Glu
                680                 685                 690
Asn Pro Phe Ala Lys Ala Asn Gly Leu Pro Gly Lys Leu Met Pro
                695                 700                 705
Ala Pro Leu Leu Ser Leu Pro Glu Asp Lys Ile Leu Glu Thr Phe
                710                 715                 720
Arg Leu Ser Asn Lys Leu Lys Gly Ser Thr Asp Glu Met Ser Ala
                725                 730                 735
Pro Gly Ala Glu Arg Gly Pro Pro Glu Asp Arg Asp Ala Glu Pro
                740                 745                 750
Gln Pro Gly Ser Pro Ala Ala Glu Ser Leu Glu Glu Pro Asp Ala
                755                 760                 765
Ala Ala Ser Leu Phe Pro Phe Ser Glu Gly
                770                 775
```

<210> 5
<211> 351
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1873139CD1

<400> 5

```
Met Asn Ala Ile Leu Gln Ser Leu Ser Asn Ile Glu Gln Phe Cys
  1                 5                  10                  15
Cys Tyr Phe Lys Glu Leu Pro Ala Val Glu Leu Arg Asn Gly Lys
                 20                  25                  30
Thr Ala Gly Arg Arg Thr Tyr His Thr Arg Ser Gln Gly Asp Asn
                 35                  40                  45
```

```
Asn Val Ser Leu Val Glu Glu Phe Arg Lys Thr Leu Cys Ala Leu
                  50                  55                  60
Trp Gln Gly Ser Gln Thr Ala Phe Ser Pro Glu Ser Leu Phe Tyr
                  65                  70                  75
Val Val Trp Lys Ile Met Pro Asn Phe Arg Gly Tyr Gln Gln Gln
                  80                  85                  90
Asp Ala His Glu Phe Met Arg Tyr Leu Leu Asp His Leu His Leu
                  95                 100                 105
Glu Leu Gln Gly Gly Phe Asn Gly Val Ser Arg Ser Ala Ile Leu
                 110                 115                 120
Gln Glu Asn Ser Thr Leu Ser Ala Ser Asn Lys Cys Cys Ile Asn
                 125                 130                 135
Gly Ala Ser Thr Val Val Thr Ala Ile Phe Gly Gly Ile Leu Gln
                 140                 145                 150
Asn Glu Val Asn Cys Leu Ile Cys Gly Thr Glu Ser Arg Lys Phe
                 155                 160                 165
Asp Pro Phe Leu Asp Leu Ser Leu Asp Ile Pro Ser Gln Phe Arg
                 170                 175                 180
Ser Lys Arg Ser Lys Asn Gln Glu Asn Gly Pro Val Cys Ser Leu
                 185                 190                 195
Arg Asp Cys Leu Arg Ser Phe Thr Asp Leu Glu Glu Leu Asp Glu
                 200                 205                 210
Thr Glu Leu Tyr Met Cys His Lys Cys Lys Lys Lys Gln Lys Ser
                 215                 220                 225
Thr Lys Lys Phe Trp Ile Gln Lys Leu Pro Lys Val Leu Cys Leu
                 230                 235                 240
His Leu Lys Arg Phe His Trp Thr Ala Tyr Leu Arg Asn Lys Val
                 245                 250                 255
Asp Thr Tyr Val Glu Phe Pro Leu Arg Gly Leu Asp Met Lys Cys
                 260                 265                 270
Tyr Leu Leu Glu Pro Glu Asn Ser Gly Pro Glu Ser Cys Leu Tyr
                 275                 280                 285
Asp Leu Ala Ala Val Val Val His His Gly Ser Gly Val Gly Ser
                 290                 295                 300
Gly His Tyr Thr Ala Tyr Ala Thr His Glu Gly Arg Trp Phe His
                 305                 310                 315
Phe Asn Asp Ser Thr Val Thr Leu Thr Asp Glu Glu Thr Val Val
                 320                 325                 330
Lys Ala Lys Ala Tyr Ile Leu Phe Tyr Val Glu His Gln Ala Lys
                 335                 340                 345
Ala Gly Ser Asp Lys Leu
                 350
```

<210> 6
<211> 136
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1903112CD1

```
<400> 6
Met Ala Leu Met Gln Arg Ser Asp Ile Phe Arg Val Ala Ile Ala
  1                 5                  10                  15
Gly Ala Pro Val Thr Leu Trp Ile Phe Tyr Asp Thr Gly Tyr Thr
                 20                  25                  30
Glu Arg Tyr Met Gly His Pro Asp Gln Asn Glu Gln Gly Tyr Tyr
                 35                  40                  45
Leu Gly Ser Val Ala Met Gln Ala Glu Lys Phe Pro Ser Glu Pro
                 50                  55                  60
Asn Arg Leu Leu Leu Leu His Gly Phe Leu Asp Glu Asn Val His
                 65                  70                  75
Phe Ala His Thr Ser Ile Leu Leu Ser Phe Leu Val Arg Ala Gly
                 80                  85                  90
Lys Pro Tyr Asp Leu Gln Ile Tyr Pro Gln Glu Arg His Ser Ile
                 95                 100                 105
Arg Val Pro Glu Ser Gly Glu His Tyr Glu Leu His Leu Leu His
```

```
                110                 115                 120
Tyr Leu Gln Glu Asn Leu Gly Ser Arg Ile Ala Ala Leu Lys Val
            125                 130                 135
Ile

<210> 7
<211> 396
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1993044CD1

<400> 7
Met Ser Leu Gly Trp Leu Glu Arg Pro Pro Ala Leu Ser Arg Ala
  1               5                  10                  15
Ala Gly Asp Gly Ala Arg Arg Leu Ser Gly Ser Met Arg Gly Asp
            20                  25                  30
Val Trp Leu Thr Ser Ser Ala Ala Gly Leu Leu Arg Ser Val Ala
        35                  40                  45
Gly Gly Ser Trp Cys Gly Gly Gln Leu Arg Ala Arg Gly Gly Ser
    50                  55                  60
Gly Arg Cys Val Ala Arg Ala Met Thr Gly Asn Ala Gly Glu Trp
65                  70                  75
Cys Leu Met Glu Ser Asp Pro Gly Val Phe Thr Glu Leu Ile Lys
                80                  85                  90
Gly Phe Gly Cys Arg Gly Ala Gln Val Glu Glu Ile Trp Ser Leu
            95                 100                 105
Glu Pro Glu Asn Phe Glu Lys Leu Lys Pro Val His Gly Leu Ile
        110                 115                 120
Phe Leu Phe Lys Trp Gln Pro Gly Glu Glu Pro Ala Gly Ser Val
    125                 130                 135
Val Gln Asp Ser Arg Leu Asp Thr Ile Phe Phe Ala Lys Gln Val
    140                 145                 150
Ile Asn Asn Ala Cys Ala Thr Gln Ala Ile Val Ser Val Leu Leu
    155                 160                 165
Asn Cys Thr His Gln Asp Val His Leu Gly Glu Thr Leu Ser Glu
    170                 175                 180
Phe Lys Glu Phe Ser Gln Ser Phe Asp Ala Ala Met Lys Gly Leu
    185                 190                 195
Ala Leu Ser Asn Ser Asp Val Ile Arg Gln Val His Asn Ser Phe
    200                 205                 210
Ala Arg Gln Gln Met Phe Glu Phe Asp Thr Lys Thr Ser Ala Lys
    215                 220                 225
Glu Glu Asp Ala Phe His Phe Val Ser Tyr Val Pro Val Asn Gly
    230                 235                 240
Arg Leu Tyr Glu Leu Asp Gly Leu Arg Glu Gly Pro Ile Asp Leu
    245                 250                 255
Gly Ala Cys Asn Gln Asp Asp Trp Phe Ser Ala Val Arg Pro Val
    260                 265                 270
Ile Glu Lys Arg Ile Gln Lys Tyr Ser Glu Gly Glu Ile Arg Phe
    275                 280                 285
Asn Leu Met Ala Ile Val Ser Asp Arg Lys Met Ile Tyr Glu Gln
    290                 295                 300
Lys Ile Ala Glu Leu Gln Arg Gln Leu Ala Glu Glu Glu Pro Met
305                 310                 315
Asp Thr Asp Gln Gly Asn Ser Met Leu Ser Ala Ile Gln Ser Glu
            320                 325                 330
Val Ala Lys Asn Gln Met Leu Ile Glu Glu Glu Val Gln Lys Leu
        335                 340                 345
Lys Arg Tyr Lys Ile Glu Asn Ile Arg Arg Lys His Asn Tyr Leu
    350                 355                 360
Pro Phe Ile Met Glu Leu Leu Lys Thr Leu Ala Glu His Gln Gln
365                 370                 375
Leu Ile Pro Leu Val Glu Lys Ala Lys Glu Lys Gln Asn Ala Lys
            380                 385                 390
```

Lys Ala Gln Glu Thr Lys
395

<210> 8
<211> 246
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2292182CD1

<400> 8
Met Ala Gly Ala Pro Asp Glu Arg Arg Arg Gly Pro Ala Ala Gly
1 5 10 15
Glu Gln Leu Gln Gln Gln His Val Ser Cys Gln Val Phe Pro Glu
20 25 30
Arg Leu Ala Gln Gly Asn Pro Gln Gln Gly Phe Phe Ser Ser Phe
35 40 45
Phe Thr Ser Asn Gln Lys Cys Gln Leu Arg Leu Leu Lys Thr Leu
50 55 60
Glu Thr Asn Pro Tyr Val Lys Leu Leu Leu Asp Ala Met Lys His
65 70 75
Ser Gly Cys Ala Val Asn Lys Asp Arg His Phe Ser Cys Glu Asp
80 85 90
Cys Asn Gly Asn Val Ser Gly Gly Phe Asp Ala Ser Thr Ser Gln
95 100 105
Ile Val Leu Cys Gln Asn Asn Ile His Asn Gln Ala His Met Asn
110 115 120
Arg Val Val Thr His Glu Leu Ile His Ala Phe Asp His Cys Arg
125 130 135
Ala His Val Asp Trp Phe Thr Asn Ile Arg His Leu Ala Cys Ser
140 145 150
Glu Val Arg Ala Ala Asn Leu Ser Gly Asp Cys Ser Leu Val Asn
155 160 165
Glu Ile Phe Arg Leu His Phe Gly Leu Lys Gln His His Gln Thr
170 175 180
Cys Val Arg Asp Arg Ala Thr Leu Ser Ile Leu Ala Val Arg Asn
185 190 195
Ile Ser Lys Glu Val Ala Lys Lys Ala Val Asp Glu Val Phe Glu
200 205 210
Ser Cys Phe Asn Asp His Glu Pro Phe Gly Arg Ile Pro His Asn
215 220 225
Lys Thr Tyr Ala Arg Tyr Ala His Arg Asp Phe Glu Asn Arg Asp
230 235 240
Arg Tyr Tyr Ser Asn Ile
245

<210> 9
<211> 262
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2331301CD1

<400> 9
Met Glu Val Tyr Ile Arg His Leu Glu Lys Val Leu Arg Arg Tyr
1 5 10 15
Val Gln Arg Leu Gln Trp Leu Leu Ser Gly Ser Arg Arg Leu Phe
20 25 30
Gly Thr Val Leu Glu Ser Lys Val Cys Ile Leu Leu Asp Thr Ser
35 40 45
Gly Ser Met Gly Pro Tyr Leu Gln Gln Val Lys Thr Glu Leu Val
50 55 60
Leu Leu Ile Trp Glu Gln Leu Arg Lys Cys Cys Asp Ser Phe Asn
65 70 75
Leu Leu Ser Phe Ala Glu Ser Leu Gln Ser Trp Gln Asp Thr Leu

```
                  80                  85                  90
Val Glu Thr Thr Asp Ala Ala Cys His Glu Ala Met Gln Trp Val
                  95                 100                 105
Thr His Leu Gln Ala Gln Gly Ser Thr Ser Ile Leu Gln Ala Leu
                 110                 115                 120
Leu Lys Ala Phe Ser Phe His Asp Leu Glu Gly Leu Tyr Leu Leu
                 125                 130                 135
Thr Asp Gly Lys Pro Asp Thr Ser Cys Ser Leu Val Leu Asn Glu
                 140                 145                 150
Val Gln Lys Leu Arg Glu Lys Arg Asp Val Lys Val His Thr Ile
                 155                 160                 165
Ser Leu Asn Cys Ser Asp Arg Ala Ala Val Glu Phe Leu Arg Lys
                 170                 175                 180
Leu Ala Ser Phe Thr Gly Gly Arg Tyr His Cys Pro Val Gly Glu
                 185                 190                 195
Asp Thr Leu Ser Lys Ile His Ser Leu Leu Thr Lys Gly Phe Ile
                 200                 205                 210
Asn Glu Lys Asp Arg Thr Leu Pro Pro Phe Glu Gly Asp Asp Leu
                 215                 220                 225
Arg Ile Leu Ala Gln Glu Ile Thr Lys Ala Arg Ser Phe Leu Trp
                 230                 235                 240
Gln Ala Gln Ser Phe Arg Ser Gln Leu Gln Lys Lys Asn Asp Ala
                 245                 250                 255
Glu Pro Lys Val Thr Leu Ser
                 260

<210> 10
<211> 406
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2517512CD1

<400> 10
Met Ala Ala Ala Val Arg Gln Asp Leu Ala Gln Leu Met Asn Ser
  1               5                  10                  15
Ser Gly Ser His Lys Asp Leu Ala Gly Lys Tyr Arg Gln Ile Leu
                 20                  25                  30
Glu Lys Ala Ile Gln Leu Ser Gly Ala Glu Gln Leu Glu Ala Leu
                 35                  40                  45
Lys Ala Phe Val Glu Ala Met Val Asn Glu Asn Val Ser Leu Val
                 50                  55                  60
Ile Ser Arg Gln Leu Leu Thr Asp Phe Cys Thr His Leu Pro Asn
                 65                  70                  75
Leu Pro Asp Ser Thr Ala Lys Glu Ile Tyr His Phe Thr Leu Glu
                 80                  85                  90
Lys Ile Gln Pro Arg Val Ile Ser Phe Glu Glu Gln Val Ala Ser
                 95                 100                 105
Ile Arg Gln His Leu Ala Ser Ile Tyr Glu Lys Glu Glu Asp Trp
                110                 115                 120
Arg Asn Ala Ala Gln Val Leu Val Gly Ile Pro Leu Glu Thr Gly
                125                 130                 135
Gln Lys Gln Tyr Asn Val Asp Tyr Lys Leu Glu Thr Tyr Leu Lys
                140                 145                 150
Ile Ala Arg Leu Tyr Leu Glu Asp Asp Pro Val Gln Ala Glu
                155                 160                 165
Ala Tyr Ile Asn Arg Ala Ser Leu Leu Gln Asn Glu Ser Thr Asn
                170                 175                 180
Glu Gln Leu Gln Ile His Tyr Lys Val Cys Tyr Ala Arg Val Leu
                185                 190                 195
Asp Tyr Arg Arg Lys Phe Ile Glu Ala Ala Gln Arg Tyr Asn Glu
                200                 205                 210
Leu Ser Tyr Lys Thr Ile Val His Glu Ser Glu Arg Leu Glu Ala
                215                 220                 225
Leu Lys His Ala Leu His Cys Thr Ile Leu Ala Ser Ala Gly Gln
                230                 235                 240
```

```
Gln Arg Ser Arg Met Leu Ala Thr Leu Phe Lys Asp Glu Arg Cys
                245                 250                 255
Gln Gln Leu Ala Ala Tyr Gly Ile Leu Glu Lys Met Tyr Leu Asp
                260                 265                 270
Arg Ile Ile Arg Gly Asn Gln Leu Gln Glu Phe Ala Ala Met Leu
                275                 280                 285
Met Pro His Gln Lys Ala Thr Thr Ala Asp Gly Ser Ser Ile Leu
                290                 295                 300
Asp Arg Ala Val Ile Glu His Asn Leu Leu Ser Ala Ser Lys Leu
                305                 310                 315
Tyr Asn Asn Ile Thr Phe Glu Glu Leu Gly Ala Leu Leu Glu Ile
                320                 325                 330
Pro Ala Ala Lys Ala Glu Lys Ile Ala Ser Gln Met Ile Thr Glu
                335                 340                 345
Gly Arg Met Asn Gly Phe Ile Asp Gln Ile Asp Gly Ile Val His
                350                 355                 360
Phe Glu Thr Arg Glu Ala Leu Pro Thr Trp Asp Lys Gln Ile Gln
                365                 370                 375
Ser Leu Cys Phe Gln Val Asn Asn Leu Leu Glu Lys Ile Ser Gln
                380                 385                 390
Thr Ala Pro Glu Trp Thr Ala Gln Ala Met Glu Ala Gln Met Ala
                395                 400                 405
Gln

<210> 11
<211> 172
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3489039CD1

<400> 11
Met Leu Leu Pro Asn Ile Leu Leu Thr Gly Thr Pro Gly Val Gly
  1               5                  10                  15
Lys Thr Thr Leu Gly Lys Glu Leu Ala Ser Lys Ser Gly Leu Lys
                20                  25                  30
Tyr Ile Asn Val Gly Asp Leu Ala Arg Glu Glu Gln Leu Tyr Asp
                35                  40                  45
Gly Tyr Asp Glu Glu Tyr Asp Cys Pro Ile Leu Asp Glu Asp Arg
                50                  55                  60
Val Val Asp Glu Leu Asp Asn Gln Met Arg Glu Gly Gly Val Ile
                65                  70                  75
Val Asp Tyr His Gly Cys Asp Phe Phe Pro Glu Arg Trp Phe His
                80                  85                  90
Ile Val Phe Val Leu Arg Thr Asp Thr Asn Val Leu Tyr Glu Arg
                95                 100                 105
Leu Glu Thr Arg Gly Tyr Asn Glu Lys Lys Leu Thr Asp Asn Ile
               110                 115                 120
Gln Cys Glu Ile Phe Gln Val Leu Tyr Glu Glu Ala Thr Ala Ser
               125                 130                 135
Tyr Lys Glu Glu Ile Val His Gln Leu Pro Ser Asn Lys Pro Glu
               140                 145                 150
Glu Leu Glu Asn Asn Val Asp Gln Ile Leu Lys Trp Ile Glu Gln
               155                 160                 165
Trp Ile Lys Asp His Asn Ser
               170

<210> 12
<211> 517
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5432879CD1
```

<400> 12

```
Met Leu Ser Ser Arg Ala Glu Ala Ala Met Thr Ala Ala Asp Arg
  1               5                  10                  15
Ala Ile Gln Arg Phe Leu Arg Thr Gly Ala Ala Val Arg Tyr Lys
             20                  25                  30
Val Met Lys Asn Trp Gly Val Ile Gly Gly Ile Ala Ala Ala Leu
         35                  40                  45
Ala Ala Gly Ile Tyr Val Ile Trp Gly Pro Ile Thr Glu Arg Lys
     50                  55                  60
Lys Arg Arg Lys Gly Leu Val Pro Gly Leu Val Asn Leu Gly Asn
 65                  70                  75                  80
Thr Cys Phe Met Asn Ser Leu Leu Gln Gly Leu Ser Ala Cys Pro
                 85                  90                  95
Ala Phe Ile Arg Trp Leu Glu Glu Phe Thr Ser Gln Tyr Ser Arg
            100                 105                 110
Asp Gln Lys Glu Pro Pro Ser His Gln Tyr Leu Ser Leu Thr Leu
        115                 120                 125
Leu His Leu Leu Lys Ala Leu Ser Cys Gln Glu Val Thr Asp Asp
    130                 135                 140
Glu Val Leu Asp Ala Ser Cys Leu Leu Asp Val Leu Arg Met Tyr
145                 150                 155                 160
Arg Trp Gln Ile Ser Ser Phe Glu Glu Gln Asp Ala His Glu Leu
                165                 170                 175
Phe His Val Ile Thr Ser Ser Leu Glu Asp Glu Arg Asp Arg Gln
            180                 185                 190
Pro Arg Val Thr His Leu Phe Asp Val His Ser Leu Glu Gln Gln
        195                 200                 205
Ser Glu Ile Thr Pro Lys Gln Ile Thr Cys Arg Thr Arg Gly Ser
    210                 215                 220
Pro His Pro Thr Ser Asn His Trp Lys Ser Gln His Pro Phe His
225                 230                 235                 240
Gly Arg Leu Thr Ser Asn Met Val Cys Lys His Cys Glu His Gln
                245                 250                 255
Ser Pro Val Arg Phe Asp Thr Phe Asp Ser Leu Ser Leu Ser Ile
            260                 265                 270
Pro Ala Ala Thr Trp Gly His Pro Leu Thr Leu Asp His Cys Leu
        275                 280                 285
His His Phe Ile Ser Ser Glu Ser Val Arg Asp Val Val Cys Asp
    290                 295                 300
Asn Cys Thr Lys Ile Glu Ala Lys Gly Thr Leu Asn Gly Glu Lys
305                 310                 315                 320
Val Glu His Gln Arg Thr Thr Phe Val Lys Gln Leu Lys Leu Gly
                325                 330                 335
Lys Leu Pro Gln Cys Leu Cys Ile His Leu Gln Arg Leu Ser Trp
            340                 345                 350
Ser Ser His Gly Thr Pro Leu Lys Arg His Glu His Val Gln Phe
        355                 360                 365
Asn Glu Phe Leu Met Met Asp Ile Tyr Lys Tyr His Leu Leu Gly
    370                 375                 380
His Lys Pro Ser Gln His Asn Pro Lys Leu Asn Lys Asn Pro Gly
385                 390                 395                 400
Pro Thr Leu Glu Leu Gln Asp Gly Pro Gly Ala Pro Thr Pro Val
                405                 410                 415
Leu Asn Gln Pro Gly Ala Pro Lys Thr Gln Ile Phe Met Asn Gly
            420                 425                 430
Ala Cys Ser Pro Ser Leu Leu Pro Thr Leu Ser Ala Pro Met Pro
        435                 440                 445
Phe Pro Leu Pro Val Val Pro Asp Tyr Ser Ser Ser Thr Tyr Leu
    450                 455                 460
Phe Arg Leu Met Ala Val Val Val His His Gly Asp Met His Ser
465                 470                 475                 480
Gly His Phe Val Thr Tyr Arg Arg Ser Pro Pro Ser Ala Arg Asn
                485                 490                 495
Pro Leu Ser Thr Ser Asn Gln Trp Leu Trp Val Ser Asp Asp Thr
            500                 505                 510
Val Arg Lys Ala Ser Leu Gln Glu Val Leu Ser Ser Ser Ala Tyr
        515                 520                 525
```

```
Leu Leu Phe Tyr Glu Arg Val Leu Ser Arg Met Gln His Gln Ser
                500                 505                 510
Gln Glu Cys Lys Ser Glu Glu
                515
<210> 13
<211> 346
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5853753CD1

<400> 13
Met Val Glu Lys Glu Glu Ala Gly Gly Gly Ile Ser Glu Glu Glu
  1               5                  10                  15
Ala Ala Gln Tyr Asp Arg Gln Ile Arg Leu Trp Gly Leu Glu Ala
             20                  25                  30
Gln Lys Arg Leu Arg Ala Ser Arg Val Leu Leu Val Gly Leu Lys
             35                  40                  45
Gly Leu Gly Ala Glu Ile Ala Lys Asn Leu Ile Leu Ala Gly Val
             50                  55                  60
Lys Gly Leu Thr Met Leu Asp His Glu Gln Val Thr Pro Glu Asp
             65                  70                  75
Pro Gly Ala Gln Phe Leu Ile Arg Thr Gly Ser Val Gly Arg Asn
             80                  85                  90
Arg Ala Glu Ala Ser Leu Glu Arg Ala Gln Asn Leu Asn Pro Met
             95                 100                 105
Val Asp Val Lys Val Asp Thr Glu Asp Ile Glu Lys Lys Pro Glu
            110                 115                 120
Ser Phe Phe Thr Gln Phe Asp Ala Val Cys Leu Thr Cys Cys Ser
            125                 130                 135
Arg Asp Val Ile Val Lys Val Asp Gln Ile Cys His Lys Asn Ser
            140                 145                 150
Ile Lys Phe Phe Thr Gly Asp Val Phe Gly Tyr His Gly Tyr Thr
            155                 160                 165
Phe Ala Asn Leu Gly Glu His Glu Phe Val Glu Glu Lys Thr Lys
            170                 175                 180
Val Ala Lys Val Ser Gln Gly Val Glu Asp Gly Pro Asp Thr Lys
            185                 190                 195
Arg Ala Lys Leu Asp Ser Ser Glu Thr Thr Met Val Lys Lys Lys
            200                 205                 210
Val Val Phe Cys Pro Val Lys Glu Ala Leu Glu Val Asp Trp Ser
            215                 220                 225
Ser Glu Lys Ala Lys Ala Ala Leu Lys Arg Thr Thr Ser Asp Tyr
            230                 235                 240
Phe Leu Leu Gln Val Leu Leu Lys Phe Arg Thr Asp Lys Gly Arg
            245                 250                 255
Asp Pro Ser Ser Asp Thr Tyr Glu Glu Asp Ser Glu Leu Leu Leu
            260                 265                 270
Gln Ile Arg Asn Asp Val Leu Asp Ser Leu Gly Ile Ser Pro Asp
            275                 280                 285
Leu Leu Pro Glu Asp Phe Val Arg Tyr Cys Phe Ser Glu Met Ala
            290                 295                 300
Pro Val Cys Ala Val Val Gly Gly Ile Leu Ala Gln Glu Ile Val
            305                 310                 315
Lys Ala Leu Ser Gln Arg Asp Pro Pro His Asn Asn Phe Phe Phe
            320                 325                 330
Phe Asp Gly Met Lys Gly Asn Gly Ile Val Glu Cys Leu Gly Pro
            335                 340                 345
Lys

<210> 14
<211> 151
<212> PRT
<213> Homo sapiens
```

<220>
<221> misc_feature
<223> Incyte ID No: 411344CD1

<400> 14
Met Ala Ser Met Gln Lys Arg Leu Gln Lys Glu Leu Leu Ala Leu
1 5 10 15
Gln Asn Asp Pro Pro Pro Gly Met Thr Leu Asn Glu Lys Ser Val
20 25 30
Gln Asn Ser Ile Thr Gln Trp Ile Val Asp Met Glu Gly Ala Pro
35 40 45
Gly Thr Leu Tyr Glu Gly Glu Lys Phe Gln Leu Leu Phe Lys Phe
50 55 60
Ser Ser Arg Tyr Pro Phe Asp Ser Pro Gln Val Met Phe Thr Gly
65 70 75
Glu Asn Ile Pro Val His Pro His Val Tyr Ser Asn Gly His Ile
80 85 90
Cys Leu Ser Ile Leu Thr Glu Asp Trp Ser Pro Ala Leu Ser Val
95 100 105
Gln Ser Val Cys Leu Ser Ile Ile Ser Met Leu Ser Ser Cys Lys
110 115 120
Glu Lys Arg Arg Pro Pro Asp Asn Ser Phe Tyr Val Arg Thr Cys
125 130 135
Asn Lys Asn Pro Lys Lys Thr Lys Trp Trp Tyr His Asp Asp Thr
140 145 150
Cys

<210> 15
<211> 362
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1256390CD1

<400> 15
Met Leu Val Pro Gly Gly Leu Gly Tyr Asp Arg Ser Leu Ala Gln
1 5 10 15
His Arg Gln Glu Ile Val Asp Lys Ser Val Ser Pro Trp Ser Leu
20 25 30
Glu Thr Tyr Ser Tyr Asn Ile Tyr His Pro Met Gly Glu Ile Tyr
35 40 45
Glu Trp Met Arg Glu Ile Ser Glu Lys Tyr Lys Glu Val Val Thr
50 55 60
Gln His Phe Leu Gly Val Thr Tyr Glu Thr His Pro Met Tyr Tyr
65 70 75
Leu Lys Ile Ser Gln Pro Ser Gly Asn Pro Lys Lys Ile Ile Trp
80 85 90
Met Asp Cys Gly Ile His Ala Arg Glu Trp Ile Ala Pro Ala Phe
95 100 105
Cys Gln Trp Phe Val Lys Glu Ile Leu Gln Asn His Lys Asp Asn
110 115 120
Ser Ser Ile Arg Lys Leu Leu Arg Asn Leu Asp Phe Tyr Val Leu
125 130 135
Pro Val Leu Asn Ile Asp Gly Tyr Ile Tyr Thr Trp Thr Thr Asp
140 145 150
Arg Leu Trp Arg Lys Ser Arg Ser Pro His Asn Asn Gly Thr Cys
155 160 165
Phe Gly Thr Asp Leu Asn Arg Asn Phe Asn Ala Ser Trp Cys Ser
170 175 180
Ile Gly Ala Ser Arg Asn Cys Gln Asp Gln Thr Phe Cys Gly Thr
185 190 195
Gly Pro Val Ser Glu Pro Glu Thr Lys Ala Val Ala Ser Phe Ile
200 205 210
Glu Ser Lys Lys Asp Asp Ile Leu Cys Phe Leu Thr Met His Ser
215 220 225

```
Tyr Gly Gln Leu Ile Leu Thr Pro Tyr Gly Tyr Thr Lys Asn Lys
                230                 235                 240
Ser Ser Asn His Pro Glu Met Ile Gln Val Gly Gln Lys Ala Ala
                245                 250                 255
Asn Ala Leu Lys Ala Lys Tyr Gly Thr Asn Tyr Arg Val Gly Ser
                260                 265                 270
Ser Ala Asp Ile Leu Tyr Ala Ser Ser Gly Ser Ser Arg Asp Trp
                275                 280                 285
Ala Arg Asp Ile Gly Ile Pro Phe Ser Tyr Thr Phe Glu Leu Arg
                290                 295                 300
Asp Ser Gly Thr Tyr Gly Phe Val Leu Pro Glu Ala Gln Ile Gln
                305                 310                 315
Pro Thr Cys Glu Glu Thr Met Glu Ala Val Leu Ser Val Leu Asp
                320                 325                 330
Asp Val Tyr Ala Lys His Trp His Ser Asp Ser Ala Gly Arg Val
                335                 340                 345
Thr Ser Ala Thr Met Leu Leu Gly Leu Leu Val Ser Cys Met Ser
                350                 355                 360
Leu Leu

<210> 16
<211> 123
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1786774CD1

<400> 16
Met Ser Gly Glu Glu Leu Ser Glu Ser Thr Pro Glu Pro Gln Lys
  1               5                  10                  15
Glu Ile Ser Glu Ser Leu Ser Val Thr Arg Asp Gln Asp Glu Asp
                20                  25                  30
Asp Lys Ala Pro Glu Pro Thr Trp Ala Asp Asp Leu Pro Ala Thr
                35                  40                  45
Thr Ser Ser Glu Ala Thr Thr Thr Pro Arg Pro Leu Leu Ser Thr
                50                  55                  60
Pro Val Asp Gly Ala Glu Asp Pro Arg Cys Leu Glu Ala Leu Lys
                65                  70                  75
Pro Gly Asn Cys Gly Glu Tyr Val Val Arg Trp Tyr Tyr Asp Lys
                80                  85                  90
Gln Val Asn Ser Cys Ala Arg Phe Trp Phe Ser Gly Cys Asn Gly
                95                 100                 105
Ser Gly Asn Arg Phe Asn Ser Glu Lys Glu Cys Gln Glu Thr Cys
                110                115                 120
Ile Gln Gly

<210> 17
<211> 983
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1911808CD1

<400> 17
Met Ala Pro Arg Leu Gln Leu Glu Lys Ala Ala Trp Arg Trp Ala
  1               5                  10                  15
Glu Thr Val Arg Pro Glu Glu Val Ser Gln Glu His Ile Glu Thr
                20                  25                  30
Ala Tyr Arg Ile Trp Leu Glu Pro Cys Ile Arg Gly Val Cys Arg
                35                  40                  45
Arg Asn Cys Lys Gly Asn Pro Asn Cys Leu Val Gly Ile Gly Glu
                50                  55                  60
His Ile Trp Leu Gly Glu Ile Asp Glu Asn Ser Phe His Asn Ile
```

```
                    65                  70                  75
Asp Asp Pro Asn Cys Glu Arg Arg Lys Lys Asn Ser Phe Val Gly
                    80                  85                  90
Leu Thr Asn Leu Gly Ala Thr Cys Tyr Val Asn Thr Phe Leu Gln
                    95                 100                 105
Val Trp Phe Leu Asn Leu Glu Leu Arg Gln Ala Leu Tyr Leu Cys
                   110                 115                 120
Pro Ser Thr Cys Ser Asp Tyr Met Leu Gly Asp Gly Ile Gln Glu
                   125                 130                 135
Glu Lys Asp Tyr Glu Pro Gln Thr Ile Cys Glu His Leu Gln Tyr
                   140                 145                 150
Leu Phe Ala Leu Leu Gln Asn Ser Asn Arg Arg Tyr Ile Asp Pro
                   155                 160                 165
Ser Gly Phe Val Lys Ala Leu Gly Leu Asp Thr Gly Gln Gln Gln
                   170                 175                 180
Asp Ala Gln Glu Phe Ser Lys Leu Phe Met Ser Leu Leu Glu Asp
                   185                 190                 195
Thr Leu Ser Asn Gln Lys Asn Pro Asp Val Arg Asn Ile Val Gln
                   200                 205                 210
Gln Gln Phe Cys Gly Glu Tyr Ala Tyr Val Thr Val Cys Asn Gln
                   215                 220                 225
Cys Gly Arg Glu Ser Lys Leu Leu Ser Lys Phe Tyr Glu Leu Glu
                   230                 235                 240
Leu Asn Ile Gln Gly His Lys Gln Leu Thr Asp Cys Ile Ser Glu
                   245                 250                 255
Phe Leu Lys Glu Glu Lys Leu Glu Gly Asp Asn Arg Tyr Phe Cys
                   260                 265                 270
Glu Asn Cys Gln Ser Lys Gln Asn Ala Thr Arg Lys Ile Arg Leu
                   275                 280                 285
Leu Ser Leu Pro Cys Thr Leu Asn Leu Gln Leu Met Arg Phe Val
                   290                 295                 300
Phe Asp Arg Gln Thr Gly His Lys Lys Lys Leu Asn Thr Tyr Ile
                   305                 310                 315
Gly Phe Ser Glu Ile Leu Asp Met Glu Pro Tyr Val Glu His Lys
                   320                 325                 330
Gly Gly Ser Tyr Val Tyr Glu Leu Ser Ala Val Leu Ile His Arg
                   335                 340                 345
Gly Val Ser Ala Tyr Ser Gly His Tyr Ile Ala His Val Lys Asp
                   350                 355                 360
Pro Gln Ser Gly Glu Trp Tyr Lys Phe Asn Asp Glu Asp Ile Glu
                   365                 370                 375
Lys Met Glu Gly Lys Lys Leu Gln Leu Gly Ile Glu Glu Asp Leu
                   380                 385                 390
Ala Glu Pro Ser Lys Ser Gln Thr Arg Lys Pro Lys Cys Gly Lys
                   395                 400                 405
Gly Thr His Cys Ser Arg Asn Ala Tyr Met Leu Val Tyr Arg Leu
                   410                 415                 420
Gln Thr Gln Glu Lys Pro Asn Thr Thr Val Gln Val Pro Ala Phe
                   425                 430                 435
Leu Gln Glu Leu Val Asp Arg Asp Asn Ser Lys Phe Glu Glu Trp
                   440                 445                 450
Cys Ile Glu Met Ala Glu Met Arg Lys Gln Ser Val Asp Lys Gly
                   455                 460                 465
Lys Ala Lys His Glu Glu Val Lys Glu Leu Tyr Gln Arg Leu Pro
                   470                 475                 480
Ala Gly Ala Glu Pro Tyr Glu Phe Val Ser Leu Glu Trp Leu Gln
                   485                 490                 495
Lys Trp Leu Asp Glu Ser Thr Pro Thr Lys Pro Ile Asp Asn His
                   500                 505                 510
Ala Cys Leu Cys Ser His Asp Lys Leu His Pro Asp Lys Ile Ser
                   515                 520                 525
Ile Met Lys Arg Ile Ser Glu Tyr Ala Ala Asp Ile Phe Tyr Ser
                   530                 535                 540
Arg Tyr Gly Gly Gly Pro Arg Leu Thr Val Lys Ala Leu Cys Lys
                   545                 550                 555
Glu Cys Val Val Glu Arg Cys Arg Ile Leu Arg Leu Lys Asn Gln
                   560                 565                 570
```

Leu Asn Glu Asp Tyr Lys Thr Val Asn Asn Leu Leu Lys Ala Ala
575 580 585
Val Lys Gly Ser Asp Gly Phe Trp Val Gly Lys Ser Ser Leu Arg
590 595 600
Ser Trp Arg Gln Leu Ala Leu Glu Gln Leu Asp Glu Gln Asp Gly
605 610 615
Asp Ala Glu Gln Ser Asn Gly Lys Met Asn Gly Ser Thr Leu Asn
620 625 630
Lys Asp Glu Ser Lys Glu Glu Arg Lys Glu Glu Glu Glu Leu Asn
635 640 645
Phe Asn Glu Asp Ile Leu Cys Pro His Gly Glu Leu Cys Ile Ser
650 655 660
Glu Asn Glu Arg Arg Leu Val Ser Lys Glu Ala Trp Ser Lys Leu
665 670 675
Gln Gln Tyr Phe Pro Lys Ala Pro Glu Phe Pro Ser Tyr Lys Glu
680 685 690
Cys Cys Ser Gln Cys Lys Ile Leu Glu Arg Glu Gly Glu Glu Asn
695 700 705
Glu Ala Leu His Lys Met Ile Ala Asn Glu Gln Lys Thr Ser Leu
710 715 720
Pro Asn Leu Phe Gln Asp Lys Asn Arg Pro Cys Leu Ser Asn Trp
725 730 735
Pro Glu Asp Thr Asp Val Leu Tyr Ile Val Ser Gln Phe Phe Val
740 745 750
Glu Glu Trp Arg Lys Phe Val Arg Lys Pro Thr Arg Cys Ser Pro
755 760 765
Val Ser Ser Val Gly Asn Ser Ala Leu Leu Cys Pro His Gly Gly
770 775 780
Leu Met Phe Thr Phe Ala Ser Met Thr Lys Glu Asp Ser Lys Leu
785 790 795
Ile Ala Leu Ile Trp Pro Ser Glu Trp Gln Met Ile Gln Lys Leu
800 805 810
Phe Val Val Asp His Val Ile Lys Ile Thr Arg Ile Glu Val Gly
815 820 825
Asp Val Asn Pro Ser Glu Thr Gln Tyr Ile Ser Glu Pro Lys Leu
830 835 840
Cys Pro Glu Cys Arg Glu Gly Leu Leu Cys Gln Gln Gln Arg Asp
845 850 855
Leu Arg Glu Tyr Thr Gln Ala Thr Ile Tyr Val His Lys Val Val
860 865 870
Asp Asn Lys Lys Val Met Lys Asp Ser Ala Pro Glu Leu Asn Val
875 880 885
Ser Ser Ser Glu Thr Glu Glu Asp Lys Glu Glu Ala Lys Pro Asp
890 895 900
Gly Glu Lys Asp Pro Asp Phe Asn Gln Ile Met His Ala Phe Ser
905 910 915
Val Ala Pro Phe Asp Gln Asn Leu Ser Ile Asp Gly Lys Ile Leu
920 925 930
Ser Asp Asp Cys Ala Thr Leu Gly Thr Leu Gly Val Ile Pro Glu
935 940 945
Ser Val Ile Leu Leu Lys Ala Asp Glu Pro Ile Ala Asp Tyr Ala
950 955 960
Ala Met Asp Asp Val Met Gln Val Cys Met Pro Glu Glu Gly Phe
965 970 975
Lys Gly Thr Gly Leu Leu Gly His
980

<210> 18
<211> 227
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1973875CD1

<400> 18
Met Gly Asn Cys Val Gly Arg Gln Arg Arg Glu Arg Pro Ala Ala

```
  1               5                  10                  15
Pro Gly His Pro Arg Lys Arg Ala Gly Arg Asn Glu Pro Leu Lys
            20                  25                  30
Lys Glu Arg Leu Lys Trp Lys Ser Asp Tyr Pro Met Thr Asp Gly
        35                  40                  45
Gln Leu Arg Ser Lys Arg Asp Glu Phe Trp Asp Thr Ala Pro Ala
    50                  55                  60
Phe Glu Gly Arg Lys Glu Ile Trp Asp Ala Leu Lys Ala Ala Ala
65                  70                  75                  80
Tyr Ala Ala Glu Ala Asn Asp His Glu Leu Ala Gln Ala Ile Leu
                85                  90                  95
Asp Gly Ala Ser Ile Thr Leu Pro His Gly Thr Leu Cys Glu Cys
            100                 105                 110
Tyr Asp Glu Leu Gly Asn Arg Tyr Gln Leu Pro Ile Tyr Cys Leu
        115                 120                 125
Ser Pro Pro Val Asn Leu Leu Leu Glu His Thr Glu Glu Glu Ser
    130                 135                 140
Leu Glu Pro Pro Glu Pro Pro Pro Ser Val Arg Arg Glu Phe Pro
145                 150                 155                 160
Leu Lys Val Arg Leu Ser Thr Gly Lys Asp Val Arg Leu Ser Ala
                165                 170                 175
Ser Leu Pro Asp Thr Val Gly Gln Leu Lys Arg Gln Leu His Ala
            180                 185                 190
Gln Glu Gly Ile Glu Pro Ser Trp Gln Arg Trp Phe Phe Ser Gly
        195                 200                 205
Lys Leu Leu Thr Asp Arg Thr Arg Leu Gln Glu Thr Lys Ile Gln
    210                 215                 220
Lys Asp Phe Val Ile Gln Val Ile Ile Asn Gln Pro Pro Pro Pro
225                 230                 235                 240
Gln Asp
```

<210> 19
<211> 403
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2323917CD1

<400> 19

```
Met Glu Lys Ser Gln Lys Ile Asn Pro Phe Ile Leu His Ile Leu
  1               5                  10                  15
Gln Glu Val Asp Glu Glu Ile Lys Lys Gly Leu Ala Ala Gly Ile
            20                  25                  30
Thr Leu Asn Ile Ala Gly Asn Asn Arg Leu Val Pro Val Glu Arg
        35                  40                  45
Val Thr Gly Glu Asp Phe Trp Ile Leu Ser Lys Ile Leu Lys Asn
    50                  55                  60
Cys Leu Tyr Ile Asn Gly Leu Asp Val Gly Tyr Asn Leu Leu Cys
65                  70                  75                  80
Asp Val Gly Ala Tyr Tyr Ala Ala Lys Leu Leu Gln Lys Gln Leu
                85                  90                  95
Asn Leu Ile Tyr Leu Asn Leu Met Phe Asn Asp Ile Gly Pro Glu
            100                 105                 110
Gly Gly Glu Leu Ile Ala Lys Val Leu His Lys Asn Arg Thr Leu
        115                 120                 125
Lys Tyr Leu Arg Met Thr Gly Asn Lys Ile Glu Asn Lys Gly Gly
    130                 135                 140
Met Phe Phe Ala Ala Met Leu Gln Ile Asn Ser Ser Leu Glu Lys
145                 150                 155                 160
Leu Asp Leu Gly Asp Cys Asp Leu Gly Met Gln Ser Val Ile Ala
                165                 170                 175
Phe Ala Thr Val Leu Thr Gln Asn Gln Ala Ile Lys Ala Ile Asn
            180                 185                 190
Leu Asn Arg Pro Ile Leu Tyr Gly Glu Gln Glu Glu Ser Thr Val
        195                 200                 205
```

```
His Val Gly Leu Met Leu Lys Glu Asn His Cys Leu Val Ala Leu
                200                 205                 210
His Met Cys Lys His Asp Ile Lys Asn Ser Gly Ile Gln Gln Leu
                215                 220                 225
Cys Asp Ala Leu Tyr Leu Asn Ser Ser Leu Arg Tyr Leu Asp Val
                230                 235                 240
Ser Cys Asn Lys Ile Thr His Asp Gly Met Val Tyr Leu Ala Asp
                245                 250                 255
Val Leu Lys Ser Asn Thr Thr Leu Glu Val Ile Asp Leu Ser Phe
                260                 265                 270
Asn Arg Ile Glu Asn Ala Gly Ala Asn Tyr Leu Ser Glu Thr Leu
                275                 280                 285
Thr Ser His Asn Arg Ser Leu Lys Ala Leu Ser Val Val Ser Asn
                290                 295                 300
Asn Ile Glu Gly Glu Gly Leu Val Ala Leu Ser Gln Ser Met Lys
                305                 310                 315
Thr Asn Leu Thr Phe Ser His Ile Tyr Ile Trp Gly Asn Lys Phe
                320                 325                 330
Asp Glu Ala Thr Cys Ile Ala Tyr Ser Asp Leu Ile Gln Met Gly
                335                 340                 345
Cys Leu Lys Pro Asp Asn Thr Asp Val Glu Pro Phe Val Val Asp
                350                 355                 360
Gly Arg Val Tyr Leu Ala Glu Val Ser Asn Gly Leu Lys Lys His
                365                 370                 375
Tyr Tyr Trp Thr Ser Thr Tyr Gly Glu Ser Tyr Asp His Ser Ser
                380                 385                 390
Asn Ala Gly Phe Ala Leu Val Pro Val Gly Gln Gln Pro
                395                 400
<210> 20
<211> 372
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2754960CD1

<400> 20
Met Ser Lys Ala Phe Gly Leu Leu Arg Gln Ile Cys Gln Ser Ile
 1               5                  10                  15
Leu Ala Glu Ser Ser Gln Ser Pro Ala Asp Leu Glu Glu Lys Lys
            20                  25                  30
Glu Glu Asp Ser Asn Met Lys Arg Glu Gln Pro Arg Glu Arg Pro
        35                  40                  45
Arg Ala Trp Asp Tyr Pro His Gly Leu Val Gly Leu His Asn Ile
    50                  55                  60
Gly Gln Thr Cys Cys Leu Asn Ser Leu Ile Gln Val Phe Val Met
65                  70                  75                  80
Asn Val Asp Phe Thr Arg Ile Leu Lys Arg Ile Thr Val Pro Arg
                85                  90                  95
Gly Ala Asp Glu Gln Arg Arg Ser Val Pro Phe Gln Met Leu Leu
            100                 105                 110
Leu Leu Glu Lys Met Gln Asp Ser Arg Gln Lys Ala Val Arg Pro
        115                 120                 125
Leu Glu Leu Ala Tyr Cys Leu Gln Lys Cys Asn Val Pro Leu Phe
    130                 135                 140
Val Gln His Asp Ala Ala Gln Leu Tyr Leu Lys Leu Trp Asn Leu
145                 150                 155                 160
Ile Lys Asp Gln Ile Thr Asp Val His Leu Val Glu Arg Leu Gln
                165                 170                 175
Ala Leu Tyr Thr Ile Arg Val Lys Asp Ser Leu Ile Cys Val Asp
            180                 185                 190
Cys Ala Met Glu Ser Ser Arg Asn Ser Ser Met Leu Thr Leu Pro
        195                 200                 205
Leu Ser Leu Phe Asp Val Asp Ser Lys Pro Leu Lys Thr Leu Glu
    210                 215                 220
Asp Ala Leu His Cys Phe Phe Gln Pro Arg Glu Leu Ser Ser Lys
```

```
                215                 220                 225
Ser Lys Cys Phe Cys Glu Asn Cys Gly Lys Lys Thr Arg Gly Lys
                230                 235                 240
Gln Val Leu Lys Leu Thr His Leu Pro Gln Thr Leu Thr Ile His
                245                 250                 255
Leu Met Arg Phe Ser Ile Arg Asn Ser Gln Thr Arg Lys Ile Cys
                260                 265                 270
His Ser Leu Tyr Phe Pro Gln Ser Leu Asp Phe Ser Gln Ile Leu
                275                 280                 285
Pro Met Lys Arg Glu Ser Cys Asp Ala Glu Glu Gln Ser Gly Gly
                290                 295                 300
Gln Tyr Glu Leu Phe Ala Val Ile Ala His Val Gly Met Ala Asp
                305                 310                 315
Ser Gly His Tyr Cys Val Tyr Ile Arg Asn Ala Val Asp Gly Lys
                320                 325                 330
Trp Phe Cys Phe Asn Asp Ser Asn Ile Cys Leu Val Ser Trp Glu
                335                 340                 345
Asp Ile Gln Cys Thr Tyr Gly Asn Pro Asn Tyr His Trp Gln Glu
                350                 355                 360
Thr Ala Tyr Leu Leu Val Tyr Met Lys Met Glu Cys
                365                 370
<210> 21
<211> 94
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3092341CD1

<400> 21
Met Leu Arg Gly Val Leu Gly Lys Thr Phe Arg Leu Val Gly Tyr
  1               5                  10                  15
Thr Ile Gln Tyr Gly Cys Ile Ala His Cys Ala Phe Glu Tyr Val
                20                  25                  30
Gly Gly Val Val Met Val Pro Met Gly His Val Trp Leu Glu Gly
                35                  40                  45
Asp Asn Leu Gln Asn Ser Thr Asp Ser Arg Cys Tyr Gly Pro Ile
                50                  55                  60
Pro Tyr Gly Leu Ile Arg Gly Arg Ile Phe Phe Lys Ile Trp Leu
                65                  70                  75
Leu Ser Asp Phe Gly Phe Leu Arg Ala Ser Pro Asn Gly His Arg
                80                  85                  90
Phe Ser Asp Asp

<210> 22
<211> 248
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3658034CD1

<400> 22
Met Asn Thr Glu Arg Thr Asn Ile Gln Val Thr Val Thr Gly Pro
  1               5                  10                  15
Ser Ser Pro Ser Pro Val Lys Phe Leu Ile Asp Thr His Asn Arg
                20                  25                  30
Leu Leu Leu Gln Thr Ala Glu Leu Ala Val Val Gln Pro Thr Ala
                35                  40                  45
Val Asn Ile Ser Ala Asn Gly Phe Gly Phe Ala Ile Cys Gln Leu
                50                  55                  60
Asn Val Val Tyr Asn Val Lys Ala Ser Gly Ser Ser Arg Arg Arg
                65                  70                  75
Arg Ser Ile Gln Asn Gln Glu Ala Phe Asp Leu Asp Val Ala Val
                80                  85                  90
```

```
Lys Glu Asn Lys Asp Asp Leu Asn His Val Asp Leu Asn Val Cys
                 95                  100                 105
Thr Ser Phe Ser Gly Pro Gly Arg Ser Gly Met Ala Leu Met Glu
                110                 115                 120
Val Asn Leu Leu Ser Gly Phe Met Val Pro Ser Glu Ala Ile Ser
                125                 130                 135
Leu Ser Glu Thr Val Lys Lys Val Glu Tyr Asp His Gly Lys Leu
                140                 145                 150
Asn Leu Tyr Leu Asp Ser Val Asn Glu Thr Gln Phe Cys Val Asn
                155                 160                 165
Ile Pro Ala Val Arg Asn Phe Lys Val Ser Asn Thr Gln Asp Ala
                170                 175                 180
Ser Val Ser Ile Val Asp Tyr Tyr Glu Pro Arg Arg Gln Ala Val
                185                 190                 195
Arg Ser Tyr Asn Ser Glu Val Lys Leu Ser Ser Cys Asp Leu Cys
                200                 205                 210
Ser Asp Val Gln Gly Cys Arg Pro Cys Glu Asp Gly Ala Ser Gly
                215                 220                 225
Ser His His His Ser Ser Val Ile Phe Ile Phe Cys Phe Lys Leu
                230                 235                 240
Leu Tyr Phe Met Glu Leu Trp Leu
                245

<210> 23
<211> 520
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3883861CD1

<400> 23
Met Val Ala Arg Val Gly Leu Leu Leu Arg Ala Leu Gln Leu Leu
  1                 5                  10                  15
Leu Trp Gly His Leu Asp Ala Gln Pro Ala Glu Arg Gly Gly Gln
                 20                  25                  30
Glu Leu Arg Lys Glu Ala Glu Ala Phe Leu Glu Lys Tyr Gly Tyr
                 35                  40                  45
Leu Asn Glu Gln Val Pro Lys Ala Pro Thr Ser Thr Arg Phe Ser
                 50                  55                  60
Asp Ala Ile Arg Ala Phe Gln Trp Val Ser Gln Leu Pro Val Ser
                 65                  70                  75
Gly Val Leu Asp Arg Ala Thr Leu Arg Gln Met Thr Arg Pro Arg
                 80                  85                  90
Cys Gly Val Thr Asp Thr Asn Ser Tyr Ala Ala Trp Ala Glu Arg
                 95                 100                 105
Ile Ser Asp Leu Phe Ala Arg His Arg Thr Lys Met Arg Arg Lys
                110                 115                 120
Lys Arg Phe Ala Lys Gln Gly Asn Lys Trp Tyr Lys Gln His Leu
                125                 130                 135
Ser Tyr Arg Leu Val Asn Trp Pro Glu His Leu Pro Glu Pro Ala
                140                 145                 150
Val Arg Gly Ala Val Arg Ala Ala Phe Gln Leu Trp Ser Asn Val
                155                 160                 165
Ser Ala Leu Glu Phe Trp Glu Ala Pro Ala Thr Gly Pro Ala Asp
                170                 175                 180
Ile Arg Leu Thr Phe Phe Gln Gly Asp His Asn Asp Gly Leu Gly
                185                 190                 195
Asn Ala Phe Asp Gly Pro Gly Gly Ala Leu Ala His Ala Phe Leu
                200                 205                 210
Pro Arg Arg Gly Glu Ala His Phe Asp Gln Asp Glu Arg Trp Ser
                215                 220                 225
Leu Ser Arg Arg Arg Gly Arg Asn Leu Phe Val Val Leu Ala His
                230                 235                 240
Glu Ile Gly His Thr Leu Gly Leu Thr His Ser Pro Ala Pro Arg
                245                 250                 255
Ala Leu Met Ala Pro Tyr Tyr Lys Arg Leu Gly Arg Asp Ala Leu
```

```
                    260                 265                 270
Leu Ser Trp Asp Asp Val Leu Ala Val Gln Ser Leu Tyr Gly Lys
                    275                 280                 285
Pro Leu Gly Gly Ser Val Ala Val Gln Leu Pro Gly Lys Leu Phe
                    290                 295                 300
Thr Asp Phe Glu Thr Trp Asp Ser Tyr Ser Pro Gln Gly Arg Arg
                    305                 310                 315
Pro Glu Thr Gln Gly Pro Lys Tyr Cys His Ser Ser Phe Asp Ala
                    320                 325                 330
Ile Thr Val Asp Arg Gln Gln Gln Leu Tyr Ile Phe Lys Gly Ser
                    335                 340                 345
His Phe Trp Glu Val Ala Ala Asp Gly Asn Val Ser Glu Pro Arg
                    350                 355                 360
Pro Leu Gln Glu Arg Trp Val Gly Leu Pro Pro Asn Ile Glu Ala
                    365                 370                 375
Ala Ala Val Ser Leu Asn Asp Gly Asp Phe Tyr Phe Phe Lys Gly
                    380                 385                 390
Gly Arg Cys Trp Arg Phe Arg Gly Pro Lys Pro Val Trp Gly Leu
                    395                 400                 405
Pro Gln Leu Cys Arg Ala Gly Gly Leu Pro Arg His Pro Asp Ala
                    410                 415                 420
Ala Leu Phe Phe Pro Pro Leu Arg Arg Leu Ile Leu Phe Lys Gly
                    425                 430                 435
Ala Arg Tyr Tyr Val Leu Ala Arg Gly Gly Leu Gln Val Glu Pro
                    440                 445                 450
Tyr Tyr Pro Arg Ser Leu Gln Asp Trp Gly Gly Ile Pro Glu Glu
                    455                 460                 465
Val Ser Gly Ala Leu Pro Arg Pro Asp Gly Ser Ile Ile Phe Phe
                    470                 475                 480
Arg Asp Asp Arg Tyr Trp Arg Leu Asp Gln Ala Lys Leu Gln Ala
                    485                 490                 495
Thr Thr Ser Gly Arg Trp Ala Thr Glu Leu Pro Trp Met Gly Cys
                    500                 505                 510
Trp His Ala Asn Ser Gly Ser Ala Leu Phe
                    515                 520
```

<210> 24
<211> 422
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4993873CD1

<400> 24

```
Met Gly Pro Ala Trp Leu Trp Leu Leu Gly Thr Gly Ile Leu Ala
  1                 5                  10                  15
Ser Val His Cys Gln Pro Leu Leu Ala His Gly Asp Lys Ser Leu
                 20                  25                  30
Gln Gly Pro Gln Pro Pro Arg His Gln Leu Ser Glu Pro Ala Pro
                 35                  40                  45
Ala Tyr His Arg Ile Thr Pro Thr Ile Thr Asn Phe Ala Leu Arg
                 50                  55                  60
Leu Tyr Lys Glu Leu Ala Ala Asp Ala Pro Gly Asn Ile Phe Phe
                 65                  70                  75
Ser Pro Val Ser Ile Ser Thr Thr Leu Ala Leu Leu Ser Leu Gly
                 80                  85                  90
Ala Gln Ala Asn Thr Ser Ala Leu Ile Leu Glu Gly Leu Gly Phe
                 95                 100                 105
Asn Leu Thr Glu Thr Pro Glu Ala Asp Ile His Gln Gly Phe Arg
                110                 115                 120
Ser Leu Leu His Thr Leu Ala Leu Pro Ser Pro Lys Leu Glu Leu
                125                 130                 135
Lys Val Gly Asn Ser Leu Phe Leu Asp Lys Arg Leu Lys Pro Arg
                140                 145                 150
Gln His Tyr Leu Asp Ser Ile Lys Glu Leu Tyr Gly Ala Phe Ala
                155                 160                 165
```

```
Phe Ser Ala Asn Phe Thr Asp Ser Val Thr Thr Gly Arg Gln Ile
                170                 175                 180
Asn Asp Tyr Leu Arg Arg Gln Thr Tyr Gly Gln Val Val Asp Cys
            185                 190                 195
Leu Pro Glu Phe Ser Gln Asp Thr Phe Met Val Leu Ala Asn Tyr
        200                 205                 210
Ile Phe Phe Lys Ala Lys Trp Lys His Pro Phe Ser Arg Tyr Gln
    215                 220                 225
Thr Gln Lys Gln Ala Ser Phe Phe Val Asp Glu Arg Thr Ser Leu
230                 235                 240
Gln Val Pro Met Met His Gln Lys Glu Met His Arg Phe Leu Tyr
                245                 250                 255
Asp Gln Asp Leu Ala Cys Thr Val Leu Gln Ile Glu Tyr Arg Gly
            260                 265                 270
Asn Ala Leu Ala Leu Leu Val Leu Pro Asp Pro Gly Lys Met Lys
        275                 280                 285
Gln Val Glu Ala Ala Leu Gln Pro Gln Thr Leu Arg Lys Trp Gly
    290                 295                 300
Gln Leu Leu Leu Pro Ser Leu Leu Asp Leu His Leu Pro Arg Phe
305                 310                 315
Ser Ile Ser Gly Thr Tyr Asn Leu Glu Asp Ile Leu Pro Gln Ile
                320                 325                 330
Gly Leu Thr Asn Ile Leu Asn Leu Glu Ala Asp Phe Ser Gly Val
            335                 340                 345
Thr Gly Gln Leu Asn Lys Thr Ile Ser Lys Val Ser His Lys Ala
        350                 355                 360
Met Val Asp Met Ser Glu Lys Gly Thr Glu Ala Gly Ala Ala Ser
    365                 370                 375
Gly Leu Leu Ser Gln Pro Pro Ser Leu Asn Thr Met Ser Asp Pro
380                 385                 390
His Ala His Phe Asn Arg Pro Phe Leu Leu Leu Leu Trp Glu Val
                395                 400                 405
Thr Thr Gln Ser Leu Leu Phe Leu Gly Lys Val Val Asn Pro Val
            410                 415                 420
Ala Gly
```

<210> 25
<211> 114
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5208004CD1

```
<400> 25
Met Arg Trp Arg Gln Arg Ser Phe Leu Leu Arg Leu Phe Leu Gly
  1                 5                  10                  15
Ser Leu Arg Gly Gly Gln His His Pro Pro Leu Thr Leu Pro Ser
            20                  25                  30
Ala Ser Ser Leu Pro Phe Ser Thr Leu Ser Leu Leu Leu Ala Ser
        35                  40                  45
Ser Leu Ser Cys Cys Leu Val Ser Pro Cys Pro Lys Thr Pro Gly
    50                  55                  60
Ser Phe Val Leu Leu Pro Trp Pro Pro Pro Arg Arg Arg Ser Gln
65                  70                  75
Ala Pro Ser Pro Pro Arg Gly Ile His Thr Thr Gly Ser Cys Trp
                80                  85                  90
Gly Trp Gly Ser Pro Ala Gly Phe Leu Met Pro Cys Ala Gln Gly
            95                 100                 105
Ser Ala Ala Val Ile Phe Gly Leu Ser
        110
```

<210> 26
<211> 742
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5267783CD1

<400> 26

```
Met Pro Ala Gly Gly Lys Gly Ser His Pro Ser Ser Thr Pro Gln
 1               5                  10                  15
Arg Val Pro Asn Arg Leu Ile His Glu Lys Ser Pro Tyr Leu Leu
            20                  25                  30
Gln His Ala Tyr Asn Pro Val Asp Trp Tyr Pro Trp Gly Glu Glu
        35                  40                  45
Ala Phe Asp Lys Ala Arg Lys Glu Asn Lys Pro Ile Phe Leu Ser
    50                  55                  60
Val Gly Tyr Ser Thr Cys His Trp Cys His Met Met Glu Glu Glu
65                  70                  75
Ser Phe Gln Asn Glu Glu Ile Gly Arg Leu Leu Ser Glu Asp Phe
                80                  85                  90
Val Ser Val Lys Val Asp Arg Glu Glu Arg Pro Asp Val Asp Lys
            95                  100                 105
Val Tyr Met Thr Phe Val Gln Ala Thr Ser Ser Gly Gly Gly Trp
        110                 115                 120
Pro Met Asn Val Trp Leu Thr Pro Asn Leu Gln Pro Phe Val Gly
    125                 130                 135
Gly Thr Tyr Phe Pro Pro Glu Asp Gly Leu Thr Arg Val Gly Phe
140                 145                 150
Arg Thr Val Leu Leu Arg Ile Arg Glu Gln Trp Lys Gln Asn Lys
            155                 160                 165
Asn Thr Leu Leu Glu Asn Ser Gln Arg Val Thr Thr Ala Leu Leu
        170                 175                 180
Ala Arg Ser Glu Ile Ser Val Gly Asp Arg Gln Leu Pro Pro Ser
    185                 190                 195
Ala Ala Thr Val Asn Asn Arg Cys Phe Gln Gln Leu Asp Glu Gly
200                 205                 210
Tyr Asp Glu Glu Tyr Gly Gly Phe Ala Glu Ala Pro Lys Phe Pro
            215                 220                 225
Thr Pro Val Ile Leu Ser Phe Leu Phe Ser Tyr Trp Leu Ser His
        230                 235                 240
Arg Leu Thr Gln Asp Gly Ser Arg Ala Gln Gln Met Ala Leu His
    245                 250                 255
Thr Leu Lys Met Met Ala Asn Gly Gly Ile Arg Asp His Val Gly
260                 265                 270
Gln Gly Phe His Arg Tyr Ser Thr Asp Arg Gln Trp His Val Pro
            275                 280                 285
His Phe Glu Lys Met Leu Tyr Asp Gln Ala Gln Leu Ala Val Ala
        290                 295                 300
Tyr Ser Gln Ala Phe Gln Leu Ser Gly Asp Glu Phe Tyr Ser Asp
    305                 310                 315
Val Ala Lys Gly Ile Leu Gln Tyr Val Ala Arg Ser Leu Ser His
320                 325                 330
Arg Ser Gly Gly Phe Tyr Ser Ala Glu Asp Ala Asp Ser Pro Pro
            335                 340                 345
Glu Arg Gly Gln Arg Pro Lys Glu Gly Ala Tyr Tyr Val Trp Thr
        350                 355                 360
Val Lys Glu Val Gln Gln Leu Leu Pro Glu Pro Val Leu Gly Ala
    365                 370                 375
Thr Glu Pro Leu Thr Ser Gly Gln Leu Leu Met Lys His Tyr Gly
380                 385                 390
Leu Thr Glu Ala Gly Asn Ile Ser Pro Ser Gln Asp Pro Lys Gly
            395                 400                 405
Glu Leu Gln Gly Gln Asn Val Leu Thr Val Arg Tyr Ser Leu Glu
        410                 415                 420
Leu Thr Ala Ala Arg Phe Gly Leu Asp Val Glu Ala Val Arg Thr
    425                 430                 435
Leu Leu Asn Ser Gly Leu Glu Lys Leu Phe Gln Ala Arg Lys His
440                 445                 450
Arg Pro Lys Pro His Leu Asp Ser Lys Met Leu Ala Ala Trp Asn
            455                 460                 465
```

```
Gly Leu Met Val Ser Gly Tyr Ala Val Thr Gly Ala Val Leu Gly
                470                 475                 480
Gln Asp Arg Leu Ile Asn Tyr Ala Thr Asn Gly Ala Lys Phe Leu
                485                 490                 495
Lys Arg His Met Phe Asp Val Ala Ser Gly Arg Leu Met Arg Thr
                500                 505                 510
Cys Tyr Thr Gly Pro Gly Gly Thr Val Glu His Ser Asn Pro Pro
                515                 520                 525
Cys Trp Gly Phe Leu Glu Asp Tyr Ala Phe Val Val Arg Gly Leu
                530                 535                 540
Leu Asp Leu Tyr Glu Ala Ser Gln Glu Ser Ala Trp Leu Glu Trp
                545                 550                 555
Ala Leu Arg Leu Gln Asp Thr Gln Asp Arg Leu Phe Trp Asp Ser
                560                 565                 570
Gln Gly Gly Gly Tyr Phe Cys Ser Glu Ala Glu Leu Gly Ala Gly
                575                 580                 585
Leu Pro Leu Arg Leu Lys Asp Asp Gln Asp Gly Ala Glu Pro Ser
                590                 595                 600
Ala Asn Ser Val Ser Ala His Asn Leu Leu Arg Leu His Gly Phe
                605                 610                 615
Thr Gly His Lys Asp Trp Met Asp Lys Cys Val Cys Leu Leu Thr
                620                 625                 630
Ala Phe Ser Glu Arg Met Arg Arg Val Pro Val Ala Leu Pro Glu
                635                 640                 645
Met Val Arg Ala Leu Ser Ala Gln Gln Gln Thr Leu Lys Gln Ile
                650                 655                 660
Val Ile Cys Gly Asp Arg Gln Ala Lys Asp Thr Lys Ala Leu Val
                665                 670                 675
Gln Cys Val His Ser Val Tyr Ile Pro Asn Lys Val Leu Ile Leu
                680                 685                 690
Ala Asp Gly Asp Pro Ser Ser Phe Leu Ser Arg Gln Leu Pro Phe
                695                 700                 705
Leu Ser Thr Leu Arg Arg Leu Glu Asp Gln Ala Thr Ala Tyr Val
                710                 715                 720
Cys Glu Asn Gln Ala Cys Ser Val Pro Ile Thr Asp Pro Cys Glu
                725                 730                 735
Leu Arg Lys Leu Leu His Pro
                740
```

<210> 27
<211> 734
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5583922CD1

```
<400> 27
Met Trp Gly Leu Leu Leu Ala Leu Ala Ala Phe Ala Pro Ala Val
  1               5                  10                  15
Gly Pro Ala Leu Gly Ala Pro Arg Asn Ser Val Leu Gly Leu Ala
             20                  25                  30
Gln Pro Gly Thr Thr Lys Val Pro Gly Ser Thr Pro Ala Leu His
             35                  40                  45
Ser Ser Pro Ala Gln Pro Pro Ala Glu Thr Ala Asn Gly Thr Ser
             50                  55                  60
Glu Gln His Val Arg Ile Arg Val Ile Lys Lys Lys Lys Val Ile
             65                  70                  75
Met Lys Lys Arg Lys Lys Leu Thr Leu Thr Arg Pro Thr Pro Leu
             80                  85                  90
Val Thr Ala Gly Pro Leu Val Thr Pro Thr Pro Ala Gly Thr Leu
             95                 100                 105
Asp Pro Ala Glu Lys Gln Glu Thr Gly Cys Pro Pro Leu Gly Leu
            110                 115                 120
Glu Ser Leu Arg Val Ser Asp Ser Arg Leu Glu Ala Ser Ser Ser
            125                 130                 135
Gln Ser Phe Gly Leu Gly Pro His Arg Gly Arg Leu Asn Ile Gln
```

```
                140                 145                 150
Ser Gly Leu Glu Asp Gly Asp Leu Tyr Asp Gly Ala Trp Cys Ala
            155                 160                 165
Glu Glu Gln Asp Ala Asp Pro Trp Phe Gln Val Asp Ala Gly His
        170                 175                 180
Pro Thr Arg Phe Ser Gly Val Ile Thr Gln Gly Ser Asn Ser Val
    185                 190                 195
Trp Arg Tyr Asp Trp Val Thr Ser Tyr Lys Val Gln Phe Ser Asn
200                 205                 210
Asp Ser Arg Thr Trp Trp Gly Ser Arg Asn His Ser Ser Gly Met
215                 220                 225
Asp Ala Val Phe Pro Ala Asn Ser Asp Pro Glu Thr Pro Val Leu
230                 235                 240
Asn Leu Leu Pro Glu Pro Gln Val Ala Arg Phe Ile Arg Leu Leu
245                 250                 255
Pro Gln Thr Trp Leu Gln Gly Gly Ala Pro Cys Leu Arg Ala Glu
260                 265                 270
Ile Leu Ala Cys Pro Val Ser Asp Pro Asn Asp Leu Phe Leu Glu
275                 280                 285
Ala Pro Ala Ser Gly Ser Ser Asp Pro Leu Asp Phe Gln His His
290                 295                 300
Asn Tyr Lys Ala Met Arg Lys Leu Met Lys Gln Val Gln Glu Gln
305                 310                 315
Cys Pro Asn Ile Thr Arg Ile Tyr Ser Ile Gly Lys Ser Tyr Gln
320                 325                 330
Gly Leu Lys Leu Tyr Val Met Glu Met Ser Asp Lys Pro Gly Glu
335                 340                 345
His Glu Leu Gly Glu Pro Glu Val Arg Tyr Val Ala Gly Met His
350                 355                 360
Gly Asn Glu Ala Leu Gly Arg Glu Leu Leu Leu Leu Leu Met Gln
365                 370                 375
Phe Leu Cys His Glu Phe Leu Arg Gly Asn Pro Arg Val Thr Arg
380                 385                 390
Leu Leu Ser Glu Met Arg Ile His Leu Leu Pro Ser Met Asn Pro
395                 400                 405
Asp Gly Tyr Glu Ile Ala Tyr His Arg Gly Ser Glu Leu Val Gly
410                 415                 420
Trp Ala Glu Gly Arg Trp Asn Asn Gln Ser Ile Asp Leu Asn His
425                 430                 435
Asn Phe Ala Asp Leu Asn Thr Pro Leu Trp Glu Ala Gln Asp Asp
440                 445                 450
Gly Lys Val Pro His Ile Val Pro Asn His His Leu Pro Leu Pro
455                 460                 465
Thr Tyr Tyr Thr Leu Pro Asn Ala Thr Val Ala Pro Glu Thr Arg
470                 475                 480
Ala Val Ile Lys Trp Met Lys Arg Ile Pro Phe Val Leu Ser Ala
485                 490                 495
Asn Leu His Gly Gly Glu Leu Val Val Ser Tyr Pro Phe Asp Met
500                 505                 510
Thr Arg Thr Pro Trp Ala Ala Arg Glu Leu Thr Pro Thr Pro Asp
515                 520                 525
Asp Ala Val Phe Arg Trp Leu Ser Thr Val Tyr Ala Gly Ser Asn
530                 535                 540
Leu Ala Met Gln Asp Thr Ser Arg Arg Pro Cys His Ser Gln Asp
545                 550                 555
Phe Ser Val His Gly Asn Ile Ile Asn Gly Ala Asp Trp His Thr
560                 565                 570
Val Pro Gly Ser Met Asn Asp Phe Ser Tyr Leu His Thr Asn Cys
575                 580                 585
Phe Glu Val Thr Val Glu Leu Ser Cys Asp Lys Phe Pro His Glu
590                 595                 600
Asn Glu Leu Pro Gln Glu Trp Glu Asn Asn Lys Asp Ala Leu Leu
605                 610                 615
Thr Tyr Leu Glu Gln Val Arg Met Gly Ile Ala Gly Val Val Arg
620                 625                 630
Asp Lys Asp Thr Glu Leu Gly Ile Ala Asp Ala Val Ile Ala Val
635                 640                 645
```

```
Asp Gly Ile Asn His Asp Val Thr Thr Ala Trp Gly Gly Asp Tyr
                650                 655                 660
Trp Arg Leu Leu Thr Pro Gly Asp Tyr Met Val Thr Ala Ser Ala
                665                 670                 675
Glu Gly Tyr His Ser Val Thr Arg Asn Cys Arg Val Thr Phe Glu
                680                 685                 690
Glu Gly Pro Phe Pro Cys Asn Phe Val Leu Thr Lys Thr Pro Lys
                695                 700                 705
Gln Arg Leu Arg Glu Leu Leu Ala Ala Gly Ala Lys Val Pro Pro
                710                 715                 720
Asp Leu Arg Arg Arg Leu Glu Arg Leu Arg Gly Gln Lys Asp
                725                 730
```

<210> 28
<211> 2080
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 088718CB1

<400> 28

```
tgaaggactt ttccaggacc caaggccaca cactggaagt cttgcagctg aagggaggca 60
ctccttggcc tccgcagccg atcacatgaa ggtggtgcca agtctcctgc tctccgtcct 120
cctggcacag gtgtggctgg tacccggctt ggcccccagt cctcagtcgc cagagacccc 180
agcccctcag aaccagacca gcagggtagt gcaggctccc agggaggaag aggaagatga 240
gcaggaggcc agcgaggaga aggccggtga ggaagagaaa gcctggctga tggccagcag 300
gcagcagctt gccaaggaga cttcaaactt cggattcagc ctgctgcgaa agatctccat 360
gaggcacgat ggcaacatgg tcttctctcc atttggcatg tccttggcca tgacaggctt 420
gatgctgggg gccacagggc cgactgaaac ccagatcaag agagggctcc acttgcaggc 480
cctgaagccc accaagcccg ggctcctgcc ttccctcttt aagggactca gagagaccct 540
ctcccgcaac ctggaactgg gcctctcaca ggggagtttt gccttcatcc acaaggattt 600
tgatgtcaaa gagactttct tcaatttatc caagaggtat tttgatacag agtgcgtgcc 660
tatgaatttt cgcaatgcct cacaggccaa aggctcatg aatcattaca ttaacaaaga 720
gactcggggg aaaattccca aactgtttga tgagattaat cctgaaacca aattaattct 780
tgtggattac atcttgttca aagggaaatg gttgacccca tttgaccctg tcttcaccga 840
agtcgacact ttccacctgg acaagtacaa gaccattaag gtgcccatga tgtacggtgc 900
aggcaagttt gcctccacct ttgacaagaa ttttcgttgt catgtcctca aactgcccta 960
ccaaggaaat gccaccatgc tggtggtcct catggagaaa atgggtgacc acctcgccct 1020
tgaagactac ctgaccacag acttggtgga gacatggctc agaaacatga aaaccagaaa 1080
catggaagtt ttctttccga agttcaagct agatcagaag tatgagatgc atgagctgct 1140
taggcagatg ggaatcagaa gaatcttctc acccttttgct gaccttagtg aactctcagc 1200
tactggaaga aatctccaag tatccagggt tttacaaaga acagtgattg aagttgatga 1260
aaggggcact gaggcagtgg caggaatctt gtcagaaatt actgcttatt ccatgcctcc 1320
tgtcatcaaa gtggaccggc catttcattt catgatctat gaagaaacct ctggaatgct 1380
tctgtttctg ggcagggtgg tgaatccgac tctcctataa ttcaggacat gcataagcaa 1440
cttcgtgctg tagtagatgc tgaatctgag gtatcaaaca cacacaggat accagcaatg 1500
gatggcaggg gagagtgttc cttttgttct taactagttt agggtgttct caaataaata 1560
cagtagtccc cacttatctg aggggggatac attcaaagac ccccagcaga tgcctgaaac 1620
ggtggacagt gctgaacctt atatatattt tttcctacac atacatacct atgataaagt 1680
ttaatttata aattaggcac agtaagagat taacaataat aacaacatta agtaaaatga 1740
gttacttgaa cgcaagcact gcaataccat aacagtcaaa ctgattatag agaaggctac 1800
taagtgactc atgggcgagg agcatagaca gtgtggagac attgggcaag gggagaattc 1860
acatcctggg tgggacagag caggacgatg caagattcca tcccactact cagaatggca 1920
tgctgcttaa gacttttaga ttgtttattt ctggaatttt tcatttaatg tttttggacc 1980
atggttgacc atggttaact gagactgcag aaagcaaaac catggataag ggaggactac 2040
tacaaaagca ttaaattgat acatattttt taaaaaaaaa                       2080
```

<210> 29
<211> 2225
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 114551CB1

<400> 29
gtgcttggcg cctgcgctgg acgactcggc cggtagtgga gatgtccggc cggtctaagc 60
gggagtctcg cggttccact cgcgggaagc gagagtctga gtcgcgggtc agctccggtc 120
gcgtcaagcg ggagcgagat cgggagcggg agcctgaggc ggcgagctcc cggggcagcc 180
ctgtgcgcgt gaagcgggag ttcgagccgg cgagcgcgcg cgaggccccg gcttctgttg 240
tcccgtttgt gcgggtgaag cgggagcgcg aggtcgatga ggactcggag cctgagcggg 300
aggtgcgagc aaagaatggc cgagtggatt ctgaggaccg gaggagccgc cactgccgct 360
acctggacac cattaacagg agtgtgctgg actttgactt tgagaaactg tgttctatct 420
ccctctcaca catcaatgct tatgcctgtc tggtgtgtgg caagtacttt caaggccggg 480
gttttgaagtc tcacgcctac attcacagtg tccagtttag ccaccatgtt ttcctcaacc 540
tccacaccct caagttttac tgccttccag acaactatga gatcatcgat tcctcattgg 600
aggatatcac gtatgtgttg aagcccactt tcacaaagca gcaaattgca aacttggaca 660
agcaagccaa attgtcccgg gcatatgatg gtaccactta cctgccgggt attgtgggac 720
tgaataacat aaaggccaat gattatgcca acgctgtcct tcaggctcta tctaatgttc 780
ctcctctccg gaactacttt ctggaagaag acaattataa gaacatcaaa cgtcctccag 840
gggatatcat gttcttgttg gtccagcgtt ttggagagct gatgagaaag ctctggaacc 900
ctcgaaattt caaggcacat gtgtctcccc atgagatgct tcaggcagtt gtactttgca 960
gtaagaagac ttttcagatc accaaacaag gagatggcgt tgactttctg tcttggtttc 1020
tgaatgctct gcacttcagct ctgggggggca caaagaagaa aaagaagact attgtgactg 1080
atgttttcca ggggtccatg aggatcttca ctcaaaaagct tccccatcct gatctgccag 1140
cagaagaaaa agagcagttg ctccataatg acgagtacca ggagacaatg gtggagtcca 1200
cttttatgta cctgacgctg gaccttccta ctgccccgct ctacaaggac gagaaggagc 1260
agctcatcat tccccaagtg ccactcttca acatcctggc taagttcaat ggcatcactg 1320
agaaggaata taagacttac aaggagaact ttctgaagcg cttccagctt accaagttgc 1380
ctccatatct aatcttttgt atcaagagat tcactaagaa caacttcttt gttgagaaga 1440
atccaactat tgtcaatttc cctattacaa atgtggatct gagagaatac ttgtctgaag 1500
aagtacaagc agtacacaag aataccacct atgacctcat tgccaacatc gtgcatgacg 1560
gcaagccctc cgagggctcc taccggatcc acgtgcttca tcatgggaca ggcaagtggt 1620
atgaattaca agacctccag gtgactgaca tccttcccca gatgatcaca ctgtcagagg 1680
cttacattca gatttggaag aggcgagata atgatgaaac caaccagcag ggggcttgaa 1740
ggaggcgtct agggctttgc tcccaagggc tgtggctgat gatggtaaat aagaacacag 1800
aagctgtagc tgaacacagg ctggctggtg ggcttcctag gccagcccag cttgtatggg 1860
ttctggctac accagagcac caagagccca cttgcctggg atggccccac actgtcactc 1920
agctgttctt tgatcatttt tttctagatt gatgctcctt tctcccatgc attgagctcc 1980
catctagctt cagcagggca gaacccttct ccagatgtgt gtaacttatg tcttgagtat 2040
ctgggagtag ttgaagaaca gataattcct tccaaacatc aagcctggg attcttggag 2100
caagcagaaa gccagtaact tcgctctgtt agaggtggag gattttccta tggttccccc 2160
catttcctga tttgtatttt tagatggatt aaatagtctc ctgtttttaa accaaaaaaa 2220
aaaaa 2225

<210> 30
<211> 3287
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1261376CB1

<400> 30
caggcgcgtt acccggcgtg ctccgcgcgg cgccggcgaa gggacgtggg ggaagggcga 60
gggaggagga agcggtggct gctgcggatg tcggtgtgag cgagcggcgc ctgaacacac 120
ggcggctgcc gagcgcctga cccgggcctg cgccagagcc tgcaccgagc tccggggccc 180
cacacccgct acggtggccc tgcgcccgtt gctactgagg cggcgtgctc tgcattcttc 240
gctgtccagg cctgccggct ctggtgtctg ctggctcctc cttgcctcgc tgctcctcc 300
tgcttgcctg agtcaccgcc gccgccgccg ccacagccat ggccgagagt ggtgaaagcg 360
gcggtcctcc gggctcccag gatagcgccg ccggagccga aggtgctggc gccccgcgg 420
ccgctgcctc cgcggacgcc aaaatcatga aagtcaccgt gaagaccccg aaggaaaagg 480
aggaattcgc cgtgcccgag aatagctccg tccagcagtt taaggaagaa atctctaaac 540
gttttaaatc acatactgac caacttgtgt tgatatttgc tggaaaaatt ttgaaagatc 600
aagatacctt gagtcagcat ggaattcatg atggacttac tgttcacctt gtcattaaaa 660
cacaaaacag gcctcaggat cattcagctc agcaaacaaa tacagctgga agcaatgtta 720
ctacatcatc aactcctaat agtaactcta catctggttc tgctactagc aaccctttg 780
gtttaggtgg ccttggggga cttgcaggtc tgagtagctt gggtttgaat actaccaact 840
tctctgaact acagagtcag atgcagcgac aactttgtgc taaccctgaa atgatggtcc 900
agatcatgga aaatcccttt gttcagagca tgctctcaaa tcctgacctg atgagacagt 960
taattatggc caatccacaa atgcagcagt tgatacagag aaatccagaa attagtcata 1020

tgttgaataa tccagatata atgagacaaa cgttggaact tgccaggaat ccagcaatga 1080
tgcaggagat gatgaggaac caggaccgag ctttgagcaa cctagaaagc atcccagggg 1140
gatataatgc tttaaggcgc atgtacacag atattcagga accaatgctg agtgctgcac 1200
aagagcagtt tggtggtaat ccatttgctt ccttggtgag caatacatcc tctggtgaag 1260
gtagtcaacc ttcccgtaca gaaaatagag atccactacc caatccatgg gctccacaga 1320
cttcccagag ttcatcagct tccagcggca ctgccagcac tgtgggtggc actactggta 1380
gtactgccag tggcacttct gggcagagta ctactgcgcc aaatttggtg cctggagtag 1440
gagctagtat gttcaacaca ccaggaatgc agagcttgtt gcaacaaata actgaaaacc 1500
cacaacttat gcaaaacatg ttgtctgccc cctacatgag aagcatgatg cagtcactaa 1560
gccagaatcc tgaccttgct gcacagatga tgctgaataa tcccctattt gctggaaatc 1620
ctcagcttca agaacaaatg agacaacagc tcccaacttt cctccaacaa atgcagaatc 1680
ctgatacact atcagcaatg tcaaacccta gagcaatgca ggccttgtta cagattcagc 1740
agggtttaca gacattagca acggaagccc cgggcctcat cccagggttt actcctggct 1800
tgggggcatt aggaagcact ggaggctctt caggaactaa tggatctaac gccacaccta 1860
gtgaaaacac aagtcccaca gcaggaacca ctgaacctgg acatcagcag tttattcagc 1920
agatgcttca ggctcttgct ggagtaaatc ctcagctaca gaatccagaa gtcagatttc 1980
agcaacaact ggaacaactc agtgcaatgg gattttgaa ccgtgaagca aacttgcaag 2040
ctctaatagc aacaggaggt gatatcaatg cagctattga aaggttactg ggctcccagc 2100
catcatagca gcatttctgt atcttgaaaa aatgtaattt atttttgata acggctctta 2160
aactttaaaa tacctgcttt atttcatttt gactcttgga attctgtgct gttataaaca 2220
aacccaatat gatgcatttt aaggtggagt acagtaagat gtgtgggttt ttctgtattt 2280
ttcttttctg gaacagtggg aattaaggct actgcatgca tcacttctgc atttattgta 2340
atttttttaaa aacatcacct tttatagttg ggtgaccaga ttttgtcctg catctgtcca 2400
gtttatttgc tttttaaaca ttagcctatg gtagtaattt atgtagaata aaagcattaa 2460
aaagaagcaa atcatttgca ctctataatt tgtggtacag tattgcttat tgtgactttg 2520
gcatgcattt ttgcaaacaa tgctgtaaga tttatactac tgataatttc gttttatttg 2580
tatacaatat agagtatgca catttgggac tgcatttctg gaaacatact gcaataggct 2640
ctctgagcaa aacacctgta actaaaaaag tgaagataag aaaatactct taaagctgag 2700
tatttcctaa ttgtatagaa tcttacagca tctttgacaa acatctccca gcaaaagtgc 2760
cggttagtca ggtttgttga aaatacagta gaaaagctga ttctggttat ctctttaagg 2820
acaattaatt gtacagacac ataatgtaac attgtctcaa cattcattca cagattgact 2880
gtaaattacc ttaatctttg tgcagactga aggaacactg tagtataccc caaagtgcat 2940
ttgcctagga cttctcagct tctcccatag gtagtttaac aggcattaaa atttgtaatt 3000
gaaatgttgc tttcactgaa aaagtgtctt gatgtttcag ttatttttaa tcgccataaa 3060
aaaatagaac tatcttttgg gtttatctgt tttctcatgc acaggcaata cacaaattta 3120
aaatgagttg tgagccaatt gtttctgaag tgttttggta gttctattaa gaaatagtta 3180
aatattgtgc ttttcagagc ctcagagaca gggggacgcg gtgccggggt ggggcagcgg 3240
aatctgtcct ggatggggcc agcttaaata atactggcaa ccaagat 3287

<210> 31
<211> 2412
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1299481CB1

<400> 31
gggctgtgtg cggcggcggc ggcggcggcc gagggggatg gagcgagcgc cgagccgggt 60
cagagttgaa caatgaccat agttgacaaa gcttctgaat cttcagaccc atcagcctat 120
cagaatcagc ctggcagctc cgaggcagtc tcacctggag acatggatgc aggttctgcc 180
agctggggtg ctgtgtcttc attgaatgat gtgtcaaatc acacactttc tttaggacca 240
gtacctggtg ctgtagttta ttcgagttca tctgtacctg ataaatcaaa accatcacca 300
caaaaggatc aagccctagg tgatggcatc gctcctccac agaaagttct tttcccatct 360
gagaagattt gtcttaagtg gcaacaaact catagagttg gagctgggct ccagaatttg 420
ggcaatacct gttttgccaa tgcagcactg cagtgtttaa cctacacacc acctcttgcc 480
aattacatgc tatcacatga acactccaaa acatgtcatg cagaaggctt ttgtatgatg 540
tgtacaatgc aagcacatat tacccaggca ctcagtaatc ctggggacgt tattaaacca 600
atgtttgtca tcaatgagat gcggcgtata gctaggcacc tccgttttgg aaaccaagaa 660
gatgcccatg aattccttca atacactgtt gatgctatgc agaaagcatg cttgaatggc 720
agcaataaat tagacagaca cacccaggcc accactcttg tttgtcagat atttggagga 780
tacctaagat ctagagtcaa atgtttaaat tgcaagggcg tttcagatac ttttgatcca 840
tatcttgata taacattgga gataaaggct gctcagagtg tcaacaaggc attggagcag 900
tttgtgaagc cggaacagct tgatggagaa aactcgtaca agtgcagcaa gtgtaaaaag 960
atggttccag cttcaaagag gttcactatc catagatcct ctaatgttct tacactttct 1020
ctgaaacgtt ttgcaaattt taccggtgga aaaattgcta aggatgtgaa ataccctgag 1080

```
tatcttgata ttcggccata tatgtctcaa cccaacggag agccaattgt ctacgtcttg 1140
tatgcagtgc tggtccacac tggttttaat tgccatgctg gccattactt ctgctacata 1200
aaagctagca atggcctctg gtatcaaatg aatgactcca ttgtatctac tcagtatatt 1260
agatcggtac tcagccaaca agcctatgtg ctcttttata tcaggtccca tgatgtgaaa 1320
aatggaggtg aacttactca tcccacccat agccccggcc agtcctctcc ccgcccgtc 1380
atcagtcagc gggttgtcac caacaaacag gctgcgccag gctttatcgg accacagctt 1440
ccctctcaca tgataaagaa tccacctcac ttaaatggga ctggaccatt gaatgacacg 1500
ccaagcagtt ccatgtcgag tcctaacggg aattccagtg tcaacagggc tagtcctgtt 1560
aatgcttcag cttctgtcca aaactggtca gttaataggt cctcagtgat cccagaacat 1620
cctaagaaac aaaaaattac aatcagtatt cacaacaagt tgcctgttcg ccagtgtcag 1680
tctcaaccta accttcatag taattctttg gagaacccta ccaagcccgt tccctcttct 1740
accattacca attctgcagt acagtctacc tcgaacgcat ctacgatgtc agtttctagt 1800
aaagtaacaa aaccgatccc ccgcagtgaa tcctgctccc agcccgtgat gaatggcaaa 1860
tccaagctga agccgcagcgt gctggtgccc tatggcgccg agtcctctga ggactctgac 1920
gaggagtcaa aggggctggg caaggagaat gggattggta cgattgtgag ctcccactct 1980
cccggccaag atgccgaaga tgaggaggcc actccgcacg agcttcaaga acccatgacc 2040
ctaaacggtg ctaatagtgc agacagcgac agtgacccga aagaaaacgg cctagcgcct 2100
gatggtgcca gctgccaagg ccagcctgcc ctgcactcag aaaatccctt tgctaaggca 2160
aacggtcttc ctggaaagtt gatgcctgct cctttgctgt ctctcccaga agacaaaatc 2220
ttagagacct tcaggcttag caacaaactg aaaggctcga cggatgaaat gagtgcacct 2280
ggagcagaga ggggccctcc cgaggaccgc gacgccgagc ctcagcctgg cagcccccgc 2340
gccgaatccc tggaggagcc agatgcggcc gcaagcttat tcccctttag tgagggttaa 2400
ttttagcttg ca                                                     2412

<210> 32
<211> 2286
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1873139CB1

<400> 32
gaaaacccgg cgtccggaag cccgtgcttt ctttgacgca agggctcgag acgcagccgc 60
cgtcggccga gcgcccggct agaagcgaca ccagacggag cctccggagt tcctccgccc 120
ccacctcgcc gggtcctgga gccgcagtcc tcccagctgc cctcctcgtg gccatggagt 180
gtccacacct gagctccagc gtctgcattg ctccggacctc agccaagttc cccaacggct 240
ccccgtcgtc ctggtgctgc agcgttatcg ctgtgatgat tttgtggtta atgacaccaa 300
gctgggactg gtacagaaag tcagagaaca cttacagaac ttggaaaact cagctttcac 360
agctgacagg cataagaaaa gaaaactttt ggaaaactca acactaaaca gcaagttatt 420
aaaagtaaat ggaagcacca ctgccatttg tgccacaggc cttcggaatt tggggaacac 480
atgtttcatg aatgccatcc ttcagtcact cagtaacatt gagcagtttt gctgttattt 540
caaagaactg cccgccgtgg agttaaggaa tgggaaaaca gcaggaaggc ggacatacca 600
caccaggagc caaggggata acaatgtgtc tttggtagaa gagtttagaa agacactctg 660
tgctttatgg caaggcagcc agactgcatt tagcccagag tccttatttt atgttgtttg 720
gaagattatg ccaaacttta ggggctatca acagcaggac gcccatgaat tcatgcgcta 780
ccttttggac cacctacact tggaacttca gggcggttcc aacggtgttt cccgctcagc 840
aattctgcag gagaattcta ctctgtctgc aagtaacaag tgttgcataa atggagcatc 900
tactgttgtc acggctatat tcggaggcat tctccaaaat gaggttaact gcctcatatg 960
tgggacagaa tctagaaagt ttgatccatt cctagacctt tcatcaagta ttccaagtca 1020
gttcagaagt aagcgctcta agaatcaaga aaatggacca gtttgttcgt tacgagattg 1080
tcttcgcagt tttaccgact tagaagaact tgatgagaca gagttatata tgtgccataa 1140
atgcaaaaag aaacaaaagt ccacaaaaaa gttttggatt caaaaactac ccaaggtgct 1200
atgcttacat ttgaaaagat ttcattggac agcatattta agaaataaag ttgatacata 1260
cgtagaattt ccactgagag gcctagacat gaaatgctac ttactagagc ctgagaacag 1320
tggcccggag agctgcctgt atgacctcgc cgctgtggtg gtgcaccatg gttccggggt 1380
tggttctgga cattacacag catacgcaac tcacgaaggc cgctggttcc acttcaatga 1440
cagtactgta acactgactg acgaggagac tgtggtgaag gcgaaggcct acatcctttt 1500
ctacgtggaa caccaggcca aagctggatc ggataaactt taatacctcc tccaaatcat 1560
cattcaccaa ccatacagga gaaacatttc cagttttcca caaatactgg atacaagatt 1620
taatttcatt atgcactttt caatttccta ttttggattt agttttgtca atggtagtga 1680
cttactgaac atgggcacca actaattttg ttgttgttct accagaaaac ctcagcagat 1740
gttttgattt gctgctttag ttgtaataat tcaatttttа taggtagttg taagaactta 1800
gtcttatttg acttttttat tttatgttaa tgttttcagt tctcactttg aggcacattt 1860
acatcaatgc ttttgttcct ctcacatgct gaaagcaaga tgtgttcctt attgtgaaga 1920
gcgacacaac tgcctgctgc ctttccacag ctataatgga catcaggttg actctaaatc 1980
```

```
aaggatcatg tgtgcacaat acttgtggcc cacaaaattt cacaatgact gctgaggaat 2040
cattcttttt gcctgtaaaa tataacaaag ggcatcatta agtagaccag gtaattactg 2100
cttgtctctc aaggctgctg tctttatcag cactaactaa ataaatttgt tggttcagtt 2160
gtacttgtcc tgcaaataca agaattactc tctttgttgg ttttttggt tttggggcat 2220
acttgtttgc ggggaggtaa gatgggagta aagaccaaat acatgtaatg tttaaaaaaa 2280
aaaaaa                                                          2286

<210> 33
<211> 873
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1903112CB1

<400> 33
taaaaacatt tatcatcttt gaaagcactc ttacgagact tttattcatc agaatcagtg 60
aaatgttctc ccttttttga cagggtcaaa tagaaattga cgatcaggtg gaaggactcc 120
aatatctagc ttctcgatat gatttcattg acttagatcg tgtgggcatc cacggctggt 180
cctatggagg atacctctcc ctgatggcat taatgcagag gtcagatatc ttcagggttg 240
ctatttgctgg ggccccagtc actctgtgga tcttctatga tacaggatac acggaacgtt 300
atatgggtca ccctgaccag aatgaacagg gctattactt aggatctgtg gccatgcaag 360
cagaaaagtt cccctctgaa ccaaatcgtt tactgctctt acatggtttc ctggatgaga 420
atgtccattt tgcacatacc agtatattac tgagtttttt agtgagggct ggaaagccat 480
atgatttaca gatctatcct caggagagac acagcataag agttcctgaa tcgggagaac 540
attatgaact gcatcttttg cactaccttc aagaaaacct tggatcacgt attgctgctc 600
taaaagtgat ataattttga cctgtgtaga actctctggt atacactggc tatttaacca 660
aatgaggagg tttaatcaat agaaaacaca gaattgatca tcacattttg atacctgcca 720
tgtaacatct actcctgaaa ataaatgtgg tgccatgcag gggtctacgg tttgtggtag 780
taatctaata ccttaacccc acatgctcaa aatcaaatga tacatattcc tgagagaccc 840
agcaatacca taagaattac taaaaaaaaa aaa                             873

<210> 34
<211> 1658
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1993044CB1

<400> 34
ggaagagttt cgatgtctct agggtggcta gagcgtcctc ccgcgctcag tcgcgctgca 60
ggtgacggcg cccggaggct gtcgggaagt atgcggggtg acgtgtggtt gacgagctcg 120
gcggcgggtt tgctgagatc tgtggccggc ggcagctggt gcgggggcca gctgagagcg 180
agaggtggat cggggcggtg tgtggccagg gccatgacgg gcaatgccgg ggagtggtgc 240
ctcatggaaa gcgaccccgg ggtcttcacc gagctcatta aaggattcgg ttgccgagga 300
gcccaagtag aagaaatatg gagtttagag cctgagaatt ttgaaaaatt aaagccagtt 360
catgggttaa tttttctttt caagtggcag ccaggagaag aaccagcagg ctctgtggtt 420
caggactccc gacttgacac gatatttttt gctaagcagg ttaattaata tgcttgtgct 480
actcaagcca tagtgagtgt gttactgaac tgtacccacc aggatgtcca tttaggcgag 540
acattatcag agtttaaaga attttcacaa agttttgatg cagctatgaa aggcttggca 600
ctgagcaatt cagatgtgat tcgacaagta cacaacagtt tcgccagaca gcaaatgttt 660
gaatttgata cgaagacatc agcaaaagaa gaagatgctt ttcactttgt cagttatgtt 720
cctgttaatg ggagactgta tgaattagat ggattaagag aaggaccgat tgatttaggt 780
gcatgcaatc aagatgattg gttcagtgca gtaaggcctg tcatagaaaa aaggatacaa 840
aagtacagtg aaggtgaaat tcgatttaat ttaatggcca ttgtgtctga cagaaaaatg 900
atatatgagc agaagatagc agagttacaa agacaacttg cagaggagga acccatggat 960
acagatcaag gtaatagtat gttaagtgct attcagtcag aagttgccaa aaatcagatg 1020
cttattgaag aagaagtaca gaaattaaaa agatacaaga ttgagaatat cagaaggaag 1080
cataattatc tgcctttcat tatggaattg ttaaagactt tagcagaaca ccagcagtta 1140
atacccactag tagaaaaggc aaaagaaaaa cagaacgcaa agaaagctca ggaaaccaaa 1200
tgaagatgtt ttcagatatg tacacatttc tgcttctgca catattttca tggaaaccat 1260
tatgtataaa gaacttagag caacatccta attggctcag tgcacgtttg gcaatagtgc 1320
cagcctgtct tgtctttaat gcatggattc ataaacttct tccctacctg catcatgtgc 1380
atgtagtgca tattaaatga aagtgatatt aagaatgctt tcccaaattc cattatttga 1440
```

```
catttgaggt cctgaccact gtttaagttt tcctgggttg tcctaactaa ccattaatgg 1500
aggcctgaga aaatggcgcc aaacggatta ttgctttaat ctgtaaattt aaaataacgt 1560
aacaattttg cacattggtc agatccttgg attaaaccgg gttggtccta atttgctttcc 1620
tccatccata accgggaaaa aaaattccgg gttttccc                          1658

<210> 35
<211> 1255
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2292182CB1

<400> 35
gcctaacgtc ttcagcccag ccaggctgcc cttcttccct gcggagggag ggcctgggcg 60
gtcgcgttgg cgggagggag gttacctttc ccagtctcgc tctggccgcc tgagccagga 120
ggaagcagcg gcgaggtctg cgggaggcat ggcgggagct ccggacgagc gccggcgggg 180
ccccgcggca ggggagcagc tgcagcagca acacgtcttct tgccaggtct tcccccagcg 240
tctggcccag gggaatcccc agcaagggtt cttcttccagc ttcttcacca gcaaccagaa 300
gtgccagctt aggctcctga agacgctgga gacaaatcca tatgtcaaac ttctgcttga 360
tgctatgaaa cactcaggtt gtgctgttaa caaagataga cacttttctt gccgaagactg 420
taatggaaat gtcagtggag gttttgatgc ttcaacatct cagatagttt tgtgccagaa 480
taatatccat aatcaggccc atatgaacag agtggtcaca cacgagctta ttcatgcatt 540
tgatcattgt cgtgcccatg tcgactggtt caccaacatc agacatttgg cgtgctcaga 600
ggttcgagct gctaacctta gtggagactg ctcacttgtc aatgaaatat tcaggttaca 660
ttttggatta aaacaacacc accagacttg tgtgcgagac agagccactc tttctatcct 720
ggctgttagg aatatcagca aagaagtagc taaaaaggct gttgatgaag tttttgaatc 780
ttgtttcaat gaccatgaac cttttggaag gatcccacat aacaagactt atgcaagata 840
tgctcacaga gactttgaaa accgtgatcg gtattattca aatatatgag cacaatgaca 900
tttttatatt acagagcttc catcatcatg aagaaaaaaa aatttggttc tcaagtgaac 960
aaacatataa aatgtaaaca atccctacaa tccagataaa acagagaaga ctgtgattct 1020
agcatattat cagaaaaagt aactatggca aaaaatgaaa ctgctttaaa ctccaaggca 1080
gaaattgtat ctttatagct aaccatttag taaataaata cattacattt ttgtatttta 1140
gtcatttgtt tttattacat gtgattattt taaaatttga ttatgaaacc tgtgaagtgt 1200
gctcttccat gatgttgata ggtgaccact gttggtattt acattaaatg taaag       1255

<210> 36
<211> 1139
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2331301CB1

<400> 36
tgcctttaaa aaaaaaaaag gcttctaatc ctctggggaa agatctgctt ccttccccat 60
ttgatgcaat gaaataccgt gtgactcttc cccagcgttg agatccatgg agtggtgaga 120
cacatccagt ggacgcccag ggagatggag gtgtacatca ggcacttgga gaaggtgtta 180
aggcgctatg tccagaggct gcagtggctg ctgtccggga gccgccgact gtttggcacc 240
gttttggaga gcaaagtatg catattgctg gacacgtcag ggtccatggg ccectacctg 300
cagcaggtga agacagagct ggttttgctg atttgggaac agctgcggaa gtgctgtgac 360
agttttaacc tgctcagctt tgcagagagc cttcagtcat ggcaggacac gctggtggag 420
accacagatg cagcgtgtca tgaggctatg caatgggtga cccacctgca agctcagggc 480
agcacctcca tcttgcaagc attgctgaaa gctttcagtt tccatgatct ggaaggattg 540
tacctcctga ccgacggaaa gccagacaca agctgcagcc ttgtcctaaa tgaagtccaa 600
aaactcaggg agaaaagaga tgtgaaagtg cacaccattt ccttgaactg ctcagacaga 660
gcggcggttg agttcctgag aaagctggct tccttcaccg gcggacgcta tcactgccct 720
gtgggtgagg acacactctc caaaattcac agcctgctga ccaaaggctt catcaatgaa 780
aaggatcgca cattgccacc atttgaagga gatgatttaa ggatcctggc ccaggaaatc 840
accaaggcca gaagcttcct ctggcaggcc caatccttca gatcccaact ccagaagaaa 900
aatgatgcag aaccaaaggt cactctttcc tagagaagtg ttctcagaaa agtcatcctg 960
caaaggacca ctcactgagc aaatctcagc cccgagggca ggatgggatg aaatgctgtg 1020
acctgcagga aacatgattc ctggtaccag gactctctgg aagctgagga aggaaagact 1080
ttgtcttttg tgtgaggggc cattcccggg tcttaatcta actctccagc cctgtccgg  1139
```

<210> 37
<211> 1753
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2517512CB1

<400> 37
gtggcggact agcccaggct cccacacccg acgctctggc ccacacagac gctactctgt 60
agcatctcag gttccctctg gctgcactct ggaggaccac actcgttttc tttttggctg 120
ccagaggccc ccgcatccac cgctgagctg ggagaaagat ggcggcagcc gtgcgacagg 180
atttggccca gctcatgaat tcgagcggct ctcataaaga tctggctggc aagtatcgtc 240
agatcctgga aaaagccatt cagttatctg gagcagaaca actagaagct ttgaaagctt 300
ttgtggaagc aatggtaaat gagaatgtca gtctcgtgat ctcgcggcag ttgctgactg 360
atttttgcac acatcttcct aacttgcctg atagcacagc caaagaaatc tatcacttca 420
ccttggaaaa gatccagcct agagtcattt cattgagga gcaggttgct tccataagac 480
agcatcttgc atctatatat gagaaagaag aagattggag aaatgcagcc caagtgttgg 540
tgggaattcc tttggaaaca ggacaaaaac agtacaatgt agattataaa ctggagactt 600
acttgaagat tgctaggcta tatctggagg atgatgatcc agtccaggca gaggcttaca 660
taaatcgagc atcgttgctt cagaatgaat caaccaatga acaattacag atacattata 720
aggtatgcta tgcacgtgtt cttgattata gaagaaaatt cattgaagct gcacaaaggt 780
acaatgagct ctcttacaag acaatagtcc acgaaagtga aagactagag gccttaaaac 840
atgctttgca ctgtacgatc ttagcatcag cagggcagca gcgttctcgg atgctagcta 900
ctctttttaa ggatgaaagg tgccagcaac ttgctgccta tgggatccta gagaaaatgt 960
atctagatag gatcatcaga ggaaatcaac ttcaagaatt tgctgccatg ctgatgcctc 1020
accaaaaagc aactacagct gatggttcca gcatcttgga cagagctgtt attgaacaca 1080
atttgttgtc tgcaagcaaa ttatataata atattacctt cgaagaactt ggagctcttt 1140
tagagatccc tgcagctaag gcggaaaaga tagcatctca aatgataacc gaaggacgta 1200
tgaatggatt tattgaccag attgatggaa tagttcattt tgaaacacga gaagccctgc 1260
caacgtggga taagcagatc caatcacttt gtttccaagt gaataacctt ttggagaaaa 1320
ttagtcaaac agcaccagaa tggacagcac aagccatgga agcccagatg gctcagtgaa 1380
tccttgcaga acttctgtgc acatgacatc tttttccatg ttgtgcagat cagtttcact 1440
atctccaaag catttgcatc atgaccttat acatttcaat ccctttttatg ctggattccg 1500
tttaaagaag acattattag agcaggaagt acaagcattt aaaatatgta gttcccatat 1560
atttcagggt ctctgtgtat taagctaact cagatgtttt gaaagctttt tctttaaaca 1620
gaggtgaaat atctgtggct aaaaagtttg agatttgtga taacttttgta gtcatgtaaa 1680
acttaagtgc ttcatgcctc tccaaatgtg gttattctaa taaatggaga aatgagccaa 1740
aaaaaaaaaa aaa 1753

<210> 38
<211> 580
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3489039CB1

<400> 38
cgaggcgccc gaagccgtcg cggcggggac catgttgctt ccgaacatcc tgctcaccgg 60
tacaccaggg gttggaaaaa ccacactagg caaagaactt gcgtcaaaat caggactgaa 120
atacattaat gtgggtgatt tagctcgaga agagcaattg tatgatggct atgatgaaga 180
gtatgactgt cccatttttag atgaagacag agtagttgat gagttagata accaaatgag 240
agaaggtgga gttattgttg attaccatgg ttgtgatttc ttccctgaac gctggtttca 300
tatagttttt gtgctgagaa cagataccaa tgtattgtac gaaagacttg aaacaggggg 360
ttataatgag aagaaactaa cagacaatat tcagtgtgag atttttcaag ttctttatga 420
agaagccaca gcatcctaca aggaagaaat cgtgcatcag ctgcccagta ataaaccaga 480
agagctagaa aataatgtag atcagatctt gaaatggatt gagcagtgga tcaaagatca 540
taactcttga cttataaggc tagctactta ataatcactc 580

<210> 39
<211> 2186
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5432879CB1

<400> 39
gtccggcggc ggcggcggcg gtagcggagg agacggtttc aggcctccgg tgcggctgca 60
atgctgagct cccgggccga ggcggcgatg acgcggccg acagggccat ccagcgcttc 120
ctgcggaccg gggcggccgt cagatataaa gtcatgaaga actggggagt tataggtgga 180
attgctgctg ctcttgcagc aggaatatat gttatttggg gtcccattac agaaagaaag 240
aagcgtagaa aagggcttgt gcctggcctt gttaatttag ggaacacctg cttcatgaac 300
tcctgctacc aaggccctgt tgcctgtcct gctttcatca ggtggctgga agagttcacc 360
tcccagtact ccagggatca gaaggagccc ccctcacacc agtatttatc cttaacactc 420
ttgcaccttc tgaaagcctt gtcctgccaa gaagttactg atgatggagt cttagatgca 480
agctgcttgt tggatgtctt aagaatgtac agatggcaga tctcatcatt tgaagaacag 540
gatgctcacg aattattcca tgtcattacc tcgtcattgg aagatgagcg agaccgccag 600
cctcggtca cacatttgtt tgatgtgcat tccctggagc agcagtcaga aataactccc 660
aaacaaatta cctgccgcac aagagggtca cctcacccca catccaatca ctggaagtct 720
caacatcctt ttcatggaag actcactagt aatatggtct gcaaacactg tgaacaccag 780
agtcctgttc gatttgatac ctttgatagc ctttcactaa gtattccagc cgccacatgg 840
ggtcacccat tgaccctgga ccactgcctt caccacttca tctcatcaga atcagtgcgg 900
gatgttgtgt gtgacaactg tacaaagatt gaagccaagg gaacgttgaa cggggaaaag 960
gtggaacacc agaggaccac ttttgttaaa cagttaaaac tagggaagct ccctcagtgt 1020
ctctgcatcc acctacagcg gctgagctgg tccagccacg gcacgcctct gaagcggcat 1080
gagcacgtgc agttcaatga gttcctgatg atggacattt acaagtacca cctccttgga 1140
cataaaccta gtcaacacaa ccctaaactg aacaagaacc cagggcctac actggagctg 1200
caggatgggc cgggagcccc cacaccagtt ctgaatcagc cagggggccc caaaacacag 1260
atttttatga atggcgcctg ctccccatct ttattgccaa cgctgtcagc gccgatgccc 1320
ttcccctctc cagttgttcc cgactacagc tcctccacat acctcttccg gctgatggca 1380
gttgtcgtcc accatggaga catgcactct ggacactttg tcacttaccg acggtcccca 1440
ccttctgcca ggaaccctct ctcaactagc aatcagtggc tgtgggtctc cgatgacact 1500
gtccgcaagg ccagcctgca ggaggtcctg tcctccagcg cctacctgct gttctacgag 1560
cgcgtccttt ccaggatgca gcaccagagc caggagtgca agtctgaaga atgactgtgc 1620
cctcctgcaa ggctagagct gatggcactg tctgcactgt ccaggaaaaa agtaaaactg 1680
tactgttgcg tgtgcaagcg gccccactag agccttccag ccttctggtg tgttctaaga 1740
gcaggctcca cctgggagcc agccccagtt cacaccaaac caggctccct gaacagtcct 1800
gttcatgtgt gtaggtggtt ctgttgtgtt aagaaagcat tcattatgtc cggagtgtct 1860
ttttactcat ctgatacagg taatttaaaag aactcagatt cttgaagcca ccgttttcat 1920
attgtaatgt taggtgttct cagaggggag gtaccttttgt ctaatcaacg tttccactta 1980
gatcttttat ttttaataag caggcccata aaaattgttg acaagaatta atgaaattat 2040
taaaggcaac aatttagaag aaaaagtgcc tttcactttc gattgctttt gtagcacgtc 2100
cattgtgaaa tattccttcc aggctactca aaggatagca agagaacagg taaatgatgc 2160
ctaaagaaca ccttcctttt tctatg 2186

<210> 40
<211> 2522
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5853753CB1

<400> 40
ctccgggcgt gctgccggcg gcggtaggtg gcgcgcgggt ccggcgggcg gttggcttga 60
gcgggaccgg agctgaggca ggaagagccg gcgccatggt ggagaaggag gaggctggcg 120
gcggcattag cgaggaggag gcggcacagt atgaccggca gatccgcctg tggggactgg 180
aggcccagaa acggctgcgg gcctctcggg tgcttcttgt cggcttgaaa ggacttgggg 240
ctgaaattgc caagaatctc atcttggcag gagtgaaagg actgaccatg ctggatcacg 300
aacaggtaac tccagaagat cccggagctc agttcttgat tcgtactggg tctgttgccc 360
gaaatagggc tgaagcctct ttggagcgag ctcagaatct caaccccatg gtggatgtga 420
aggtggacac tgaggatata gagaagaaac cagagtcatt ttcactcaa ttcgatgctg 480
tgtgtctgac ttgctgctcc agggatgtca tagttaaagt tgaccagatc tgtcacaaaa 540
atagcatcaa gttctttaca ggagatgttt ttggctacca tggatacaca tttgccaatc 600
taggagagca tgagtttgta gaggagaaaa ctaaagttgc caaagttagc caaggagtag 660
aagatgggcc cgacaccaag agagcaaaac ttgattcttc tgagacaacg atggtcaaaa 720
agaaggtggt cttctgccct gttaaagaag ccctggaggt ggactggagc agtgagaaag 780
caaaggctgc tctgaagcgc acgacctccg actactttct ccttcaagtg ctcttaaagt 840

```
tccgtacaga taaaggaaga gatcccagtt ctgatacata tgaggaagat tctgagttgt 900
tgctccagat acgaaatgat gtgcttgact cactgggtat tagtcctgac ctgcttcctg 960
aggactttgt caggtactgc ttctccgaga tggccccagt gtgtgcggtg gttggaggga 1020
ttttggcaca ggaaattgtg aaggccctgt ctcagcggga ccctcctcac aacaacttct 1080
tcttcttcga tggcatgaag gggaatggga ttgtggagtg ccttggcccc aagtgaactc 1140
aagatttggc agccccagag atgccaactg cagcatgccc acctgtattc cctgtcccct 1200
tccttcatga aggcatctcc aggcaaggaa aactgaagtc attggcccga tacaaaacat 1260
ttcctgcaac gaaggaggtg gtgccgacgt gctgcttccc atcaccagca gctgctcgac 1320
aaggggcgca gggtggctgt ctttgttcca gcactgttca ggctgcctgt catcccgggc 1380
ctgccagctc ccctgagtga tgagcacttc caagcacccc tctgcccttt ctctgtcctt 1440
atgctgtccc ggcctcgcca gccctctggg gcattgtggg agatgcctgc caggaatgag 1500
caagctctgt tgctcgggag cctcttgtca ccttcttgga cttattcccc acctgatacc 1560
ttatagagaa aagtgtgaat tcaggtggag agtaggccca ggccccatga ggcaccagtg 1620
gaagcacagc tccaagttca gacaggtgcc cttagagagg aaaaccatga caggcaaatg 1680
catttcctct ggagtttgag accctgacaa acaacaggtg gcatctggtg tgctgttctt 1740
gagttttcgt ttaggattag ttgagttcca gctgggtttt gggagaaagg agatgctacc 1800
aagtcttgga tgttagggcg agaccctgca agttgagtat tagagagctt gtctttcaag 1860
gcaggttcct ggggcttcag ggctaggagg gaggagcctg cccttttaac agaaccccag 1920
tcacatgcgg ctcaagtcac tcagaggctg ttgcatttca gggctatgtt ggtcctttgt 1980
ttacctccta aaccacagct gtttgtgttt cacatatgtt gtgaattttc cttggttctt 2040
tttaaaggaa tgataataaa gttacttgct ttaggatttg cttgtttttc ttccacttca 2100
gaagcttctg agagggaatg ggatgatcct accagttgcc ttttcagacc tgaggctcta 2160
actcaagaga ttcctcctct ccctcaccat tcctgccacc atttttctg ggtgatgcag 2220
caagagttaa attgttcaca ttctagaatg tgtagaagct tctggctcca gttgtctaat 2280
ggataattca gctaaatagt ggctattttc agtggcaaga attataataa taaagggaag 2340
tcaaaagtga tgcttcatat gccccaagtg ccacagccgc tacgggtcag agcttttatt 2400
ggaagaccag gtttgccttg atgcagccaa agtgcgttcc agttcacccc actcttgctt 2460
gtgtttcttc tttatgtaaa ttgtgtgatg tttcaataaa tcagtctctt tcctaaaaaa 2520
aa                                                                 2522
```

<210> 41
<211> 2167
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 411344CB1

<400> 41

```
cgccggtctt ggggggatgg ttccatcatg gcgtcaatgc agaaacgact acagaaagaa 60
ctgttggctt tgcaaaatga cccacctcct ggaatgacct taaatgagaa gagtgttcaa 120
aattcaatta cacagtggat tgtagacatg gaaggtgcac caggtacctt atatgaaggg 180
gaaaaatttc aacttctatt taaatttagt agtcgatatc cttttgactc tcctcaggtc 240
atgtttactg gtgaaaatat tcctgttcat cctcatgttt atagcaatgg tcatatctgt 300
ttatccattc taacagaaga ctggtcccca gcgctctcag tccaatcagt ttgtcttagc 360
attattagca tgctttccag ctgcaaggaa aagagacgac caccggataa ttctttttat 420
gtgcgaacat gtaacaagaa tccaaagaaa acaaaatggt ggtatcatga tgatacttgt 480
tgatgccact gttatcatcc tcctagcaga agatagtcct actgagaaaa tgagcacttt 540
gatcattcag tctttgaact ttaacctttg actggaagtg acctataggc aatgaagact 600
acttcctttt actgcatttt tactcgtgtg cattctgggc gcatgttgat cgctggttca 660
gtccaggcaa ctgacatgct tttattagtc atacagtatt aatgcaggtg tcaggaaatg 720
tcaaatataa ttccattttt tatttttatt tttttaagct tttggaaaag ctccaggtcc 780
tcatgtattg tgcaataaca atgacttcct tggcggtttt ggtacgttca ttgccggcaa 840
tgggcgttgt aacaggaaaa gttttcatta actcctgcca ttcaatgatt aatgcatgat 900
agggcctatg aaatgaactt actggttata gtgggaatat aaataaagtg agggatccaa 960
cattacttta aaagtcaccc caactgttta tatttggatt ctatgcactg tgatcctaag 1020
gttaacagca tgaattaaca tgcgtcttta aaggactgta atgaaagatc attgcatatt 1080
tattgaattg tttatatcta ctgtcaagtt gttttgacat ggaagatttt caagtaacat 1140
tggcagagag gtacagtatg ttatccctat ggtgaaaata aattaatttg ttgtatatag 1200
ttcctcaatc tctgaagtaa aggtatgagt aatataggt atgaatggtt taatcaaggc 1260
tttattttgg aagtaagaaa aatggcagtg atgattaaat tgctgcagtc cataatttgg 1320
gcttgttatt tgtacattaa agattttttc caagtaagtt acactctgtt aacttcctgc 1380
tagccatcag catgagccct actgcctaaa cactatttca tttatttatg tttggaaacc 1440
ccgtaaacat ttttgtttgc aatcttgttt cttttgttat aagtcaagtt tgaatgttac 1500
aatactttta ttgaaacttt tgttaagttt tttcttgtaa attttcttta cttgtgagta 1560
tcatcttgtc ctttaatcct gtaccctaaa ataagaaata catttttgac agaggcttaa 1620
```

tgttttaaca aaagagtgtg gacatttta ttttaaaatt taggcaaaag tcactatcaa 1680
atggttgctt atttgtctca cacagccata tagtttttcc tggagggttt tgttttgttg 1740
ttgttgaaaa gactttgctt acagctagat gaaactttct atagaaaaaa aaaattgttg 1800
aaaggtccag ttctcagtac catgtgagtt aatgatacta caactaagtt ctttttaaaa 1860
agtgattaat gtatttttata aattaccttt tcacatatgc aaaatctgtt tctactacaa 1920
tgttattttt actaatgcct tattgtggca ctctttttga aacatcctgc agtgaatata 1980
tgaatcaatt tgggcttaaa actgaaagcc agttgggctg aaaggtttga aatacgtacc 2040
ccagtaaaac cattcaatcc ataattggta aataatattt taaaaatggt tttaatccgt 2100
atagatgaca ttttgtagct ttggacatgg tggtaattaa ggggcattaa ttttacactc 2160
caaggta 2167

<210> 42
<211> 1826
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1256390CB1

<220>
<221> unsure
<222> 1755
<223> a, t, c, g, or other

<400> 42
ccccgcctt tctatagcat agttctcttt aggtggaatg attcctataa gatttctcat 60
tattaaatca tgcatttttc aagatggaat caatctttga tttaatctaa gctgatattc 120
tcatttgtta gaagaacaac ctacatgcta gagagagagg aggaaatata cccacgacca 180
cacagccagt tagtatccag ttggtgctgg actccagcca ggtgtcctgc ctcatggtag 240
ttaaatgata tatagaaaag gtaaatttttt aaagaaatat ttattaatat attcctataa 300
aacattttaa aggtaaccac ataaaaatgg ttaatttttc cattccaaag taaatgctaa 360
gcatgtttat taatgaagca gtacttctga ttagtatatg acattctgaa gttaattaaa 420
ctcattgcac taaatgtgtc ttccttggta tagtggagga tttgaggatt ggaatataga 480
gtagagtgct tgcttaagcc tctgcttgaa acccttatc ttttggggat gctggttcct 540
ggagggctgg gatatgatag atccttagcc caacacagac aagagattgt ggacaagtca 600
gtgagtccat ggagcctgga gacgtattcc tataacatat accaccccat gggagagatc 660
tatgagtgga tgagagagat cagtgagaag tacaaggaag tggtgacaca gcatttccta 720
ggagtgacct atgagaccca ccccatgtat tatctgaaga tcagccaacc atctggtaat 780
cccaagaaaa tcatttggat ggactgtgga attcacgcca gagaatggat tgctcctgct 840
ttttgccaat ggttcgtcaa agaaattcta caaaaccata aagacaactc aagtatacgc 900
aagctcctta ggaacctgga cttctatgtc cttccagttc ttaacataga tggttatatc 960
tacacttgga caactgatcg tctttggagg aaatcccgtt caccccataa taatggcaca 1020
tgttttggga cggatctcaa tcgaaatttc aatgcatctt ggtgtagtat tggtgcctct 1080
agaaactgcc aagatcaaac attctgtggg acagggccag tgtctgaacc agagactaaa 1140
gctgttgcca gcttcataga gagcaagaag gatgatattt tgtgcttcct gacatgcac 1200
tcttatggc agttaattct cacacctcac ggctacacca aaaataaatc aagtaaccac 1260
ccagaaatga ttcaagttgg acagaaggca gcaaatgcat tgaaagcaaa gtatggaacc 1320
aattatagag ttggatcgag tgcagatatt ttatatgcct catcagggtc ttcaagagat 1380
tgggcccgag acattgggat tcccttctca tatacgtttg agctgaggga cagttggaaca 1440
tatgggtttg ttctgccaga agctcagatc cagcccacct gtgaggagac catggaggct 1500
gtgctgtcag tcctggatga tgtgtatgcg aaacactggc actcggacag tgctggaagg 1560
gtgacatctg ccactatgct gctgggcctg ctggtgtcct gcatgtctct tctctaagtg 1620
cattctgccc aggctttttt ctgccagaaa ttccccttta tttttcctcc tagtaaagtc 1680
tttctttgat actaatttctg acattacctg tggaattttt gctgacttcc acagattttt 1740
ttttctctcc ttctntaact gtttgtacac actttaatgt aataatttac agacttgtgt 1800
agtgagatag gatattacat gccctg 1826

<210> 43
<211> 1371
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1786774CB1

<400> 43
gcaaaaattg tttcaaaaaa tttgtgagga ttttgattcc tatctcgttc aaatttttgg 60
ttcatcgtca cctcaacctg gatttgggat gtcaggggaa gaactcagtg aatctactcc 120
agagcctcaa aaagaaattt ctgagtcatt gagtgtcacc agagaccagg atgaagatga 180
taaggctcca gagccaacgt gggctgatga tctgcctgcc actacctcat ctgaggccac 240
cacccccccc aggccactgc tcagcacccc tgtggatggg gcagaggatc ctagatgttt 300
ggaagcctgg aagcctggaa actgtggtga atatgtggtt cgatggtatt atgacaaaca 360
ggtcaacttct tgtgcccgat tttggttcag tggctgtaat ggctcaggaa atagattcaa 420
cagtgaaaag gaatgtcaag aaacctgcat tcaaggatga gcaagtaaat tggcctgtct 480
ctatcaaaag catagaagtc cctaatttcc acatattcac ccaatacaaa tacagcacta 540
tatttgagtg tatactgagt atttacaact tatacatgta attgaattct cactacagcc 600
ctaggatgta catattatta accacttata taggtaagaa agctgaggct ctgagaagtt 660
tagtaacttg tcaactgtca cccaactaaa aagtttcaga gctgaggatt tagacttaga 720
gctgtgtaac ttcaatacac agactctatc tacttcacaa cctgcaatgt gattctgatt 780
cctttaattc ctgttgtatg tactatgtca gctcaaaccc ctacccctgt ccctgcccat 840
acctccacca actcacctcc ctaacctcct tatgtccctc acagtagcaa gatgtaggtg 900
ataggaagga cttcggtgtg agaattagaa atgatgtaaa tgtttacgca ggagtgctgg 960
gataggagtc gggatggtga gggtagttag atttttgcct cacttgccct gaaagtggta 1020
ataggggaga accaatcttga attacaatta cttaaatgta tcacagactg tcactttgta 1080
ttcctccaac atgttggta acaagtgctt aatgtatgtt aaaataaaga aggttttat 1140
acccttccat taaaatatgt cagtgggccc ttccatttta tggagtggaa tgggaaggcc 1200
cttgacagcc aggaaccact tgaagctggc atccactctt gaacagagag tattaaagac 1260
aggattcaca ctgaaaagtg agccaccaaa ttgaaatgcg agtaatggag agtaacggag 1320
aatcgcttga ccctgggaag cagaggttgc agagagccgg gattgcccca g 1371

<210> 44
<211> 3634
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1911808CB1

<400> 44
gcggcggcag tggcgcgcac aggtgattga ctggccagct gcctgaagga gcgccaggtc 60
ctccttgctg gcaggtggcg aagcccattg gggcggcggt gcagaccgcg gcggcggctg 120
cggcggtctg gctcgggagg cgttcctggg gccaaggcca tggccccgcg gctgcagctg 180
gagaaggcgg cctggcgctg ggcggagacg gtgcggcccg aggaggtgtc gcaggagcat 240
atcgagaccg cttaccgcat ctggctggag ccctgcattc gcggcgtgtg cagacgaaac 300
tgcaaaggaa atccgaattg cttggttggt attggtgagc atatttggtt aggagaaata 360
gatgaaaata gttttcataa catcgatgat ccccactgtg agaggagaaa aaagaactca 420
tttgtgggcc tgactaacct tggagccact tgttatgtca acacatttct tcaagtgtgg 480
ttttctcaact tggagcttcg gcaggcactc tacttatgtc caagcacttg tagtgactac 540
atgctgggag acggcatcca agaagaaaaa gattatgagc ctcaaacaat ttgtgagcat 600
ctccagtact tgtttgcctt gttgcaaaac agtaataggc gatacattga tccatcagga 660
tttgttaaag ccttgggcct ggacactgga caacagcagg atgctcaaga attttcaaag 720
ctcttttatgt ctctattgga agatactttg tctaaccaaa agaatccaga tgtgcgcaat 780
attgttcaac agcagttctg tggagaatat gcctatgtaa ctgtttgcaa ccagtgtggc 840
agagagtcta agcttttgtc aaaatttttat gagctggagt taaatatcca aggccacaaa 900
cagttaacag attgtatctc ggaatttttg aaggaagaaa aattagaagg agacaatcgc 960
tatttttgcg agaactgtca aagcaaacag aatgcaacaa gaaagattcg acttcttagc 1020
cttccttgca ctctgaactt gcagctaatg cgttttgtct ttgacaggca aactggacat 1080
aagaaaaagc tgaatacctа cattggcttc tcagaaattt tggatatgga gccttatgtg 1140
gaacataaag gtgggtccta cgtgtatgaa ctcagcgcag tcctcataca cagaggagtg 1200
agtgcttatt ctggccacta catcgcccac gtgaaagatc cacagtctgg tgaatggtat 1260
aagtttaatg atgaagacat agaaaagatg gaggggaaga aattacaact agggattgag 1320
gaagatctag cagaaccttc taagtctcag acacgtaaac ccaagtgtgg caaaggaact 1380
cattgctctc gaaatgcata tatgttggtt tatagactgc aaactcaaga aaagcccaac 1440
actactgttc aagttccagc ctttcttcaa gagctggtag atcgggataa ttccaaattt 1500
gaggagtggt gtattgaaat ggctgagatg cgtaagcaaa gtgtggataa aggaaaagca 1560
aaacacgaag aggttaagga gctgtaccaa aggttacctg ctggagctga gccctatgag 1620
tttgtctctc tggaatggct gcaaaagtgg ttggatgaat caacacctac caaacctatt 1680
gataatcacg cttgcctgtg ttcccatgac aagcttccacc cggataaaat atcaattatg 1740
aagaggatat ctgaatatgc agctgacatt ttctatagta gatatggagg aggtccaaga 1800
ctaactgtga aagccctgtg taaggaatgt gtagtagaac gttgtcgcat attgcgtctg 1860
aagaaccaac taaatgaaga ttataaaact gttaataatc tgctgaaagc agcagtaaag 1920

```
ggcagcgatg gattttgggt ggggaagtcc tccttgcgga gttggcgcca gctagctctt 1980
gaacagctgg atgagcaaga tggtgatgca gaacaaagca acggaaagat gaacggtagc 2040
accttaaata aagatgaatc aaaggaagaa agaaaagaag aggaggaatt aaattttaat 2100
gaagatattc tgtgtccaca tggtgagtta tgcatatctg aaaatgaaag aaggcttgtt 2160
tctaaagagg cttggagcaa actgcagcag tactttccaa aggctcctga gtttccaagt 2220
tacaaagagt gctgttcaca gtgcaagatt ttagaaagag aaggggaaga aaatgaagcc 2280
ttacataaga tgattgcaaa cgagcaaaag acttctctcc caaatttgtt ccaggataaa 2340
aacagaccgt gtctcagtaa ctggccagag gatacggatg tcctctacat cgtgtctcag 2400
ttcttttgtag aagagtggcg gaaatttgtt agaaagccta caagatgcag ccctgtgtca 2460
tcagttggga acagtgctct tttgtgtccc cacgggggcc tcatgtttac atttgcttcc 2520
atgaccaaag aagattctaa acttatagct ctcatatggc ccagtgagtg gcaaatgata 2580
caaaagctct ttgttgtgga tcatgtaatt aaaatcacga gaattgaagt gggagatgta 2640
aacccttcag aaacacagta tatttctgag cccaaactct gtccagaatg cagagaaggc 2700
ttattgtgtc agcagcagag ggacctgcgt gaatacactc aagccaccat ctatgtccat 2760
aaagttgtgg ataataaaaa ggtgatgaag gattcggctc cggaactgaa tgtgagtagt 2820
tctgaaacag aggaggacaa ggaagaagct aaaccagatg gagaaaaaga tccagatttt 2880
aagattttaa gtgatgactg tgccacccta ggcaccctg gcgtcattcc tgaatctgtc 3000
atttattga aggctgatga accaattgca gattatgctg caatggatga tgtcatgcaa 3060
gtttgtatgc cagaagaagg gtttaaaggt actggtcttc ttggacatta atctttgaat 3120
acttgctgac tgctaagaaa tgaccagagg ggaagaggag tttgacatgt tagggcatta 3180
aagcaaaggt ggatttaaga attaaaccat tacatgcccc ttccaaaagg cagaaatcca 3240
ttcaaacgtg actgtcccaa atgccttatg tcaaataaag cagattgcac tgatggacat 3300
cagacttgaa ggaaatgttt ccaattttat atttaagggg ggtggtgggt ggagggggc 3360
aagtaaagac ggaacaagtt tagtagcagt aatagtaaat catgtttaca tatgagattt 3420
atagtcgtgg gaggggaata aagttctgtt atatttcctt gctcgagttt catacccagat 3480
gcgttggtcc ataaaggatt gtatcaagta gatgggacaa cattctgctc tgaacgaaaa 3540
gtaatttag agacataacc tgcttaccaa tgcctgtctt tgattcatat tctactttca 3600
ataaagcatg aaagtgaaga acttgtaaaa aaaa                              3634
```

<210> 45
<211> 1661
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1973875CB1

<400> 45

```
ctcgagccga ttcggctcga ggcggactgg tctccggccg ggcaccgtcg cgggccccc 60
tggcccggcc acctgggacc gtgctgggga gtctgccact tccctctctc ccctggcccg 120
caaagttttg gcggacccat cgctggggct gagcgcgccc ccgggggggag atcgggggagc 180
gcccgatgcc gggcggccgg agccattgac ccgggacgcc gccgtccgct gagcagccga 240
ccacccgcgc gcctccggtg catggggact ggctgaggag ccagcatggg caactgcgtg 300
gggagacagc gccgggagag gccggcagcc ccgggacacc cccgcaagcg agcaggacgc 360
aatgagcccc tgaagaaaga gcggcttaag tggaagagcg actaccccat gactgacggg 420
cagctgcgga gcaaacggga tgagttctgg gacacagcgc gtccttcga gggccgcaag 480
gagatctggg atgccctcaa ggctgccgcc tatgctgctg aagccaacga ccacgagctg 540
tacgatgagc tgggcaatcg ctaccagctg cccatctact gcctgtcacc gccggtgaac 660
ctgctgctgg agcacacgga ggaggagagc ctggagcccc ccgagcctcc acccagcgtg 720
cgccgtgagt tcccgctgaa ggtgcgcctg tccacggcca aggacgtgag gctcagcgcc 780
agcctgcccg acacagtggg gcagctcaag aggcagctgc acgcccagga gggcatcgag 840
ccatcgtggc agcggtggtt cttctccggg aagctgctca cagacgccac acggctccag 900
gagaccaaga tccagaaaga ttttgtcatc caggtcatca tcaaccagcc cccaccaccc 960
caggactgat gggcccacgg acccctggga agaggccccg cctggagcac taggccccca 1020
ccctgctgct gccttccagt gctgtcattt tcttcagggg ccctcccctc ggtgtggctg 1080
gtgggtgagc cgtgaaggga ccctgccttt cagggcacta cgcgccacca gttcccggta 1140
cccagggagc aggcagccac acacgggcct tgcaaccttg tcagagaaaa ggcgaacagg 1200
gccctcaccc tgcctgtctc ccgaagcagg ttcgagccac aagggccaac caggaggccc 1260
ctggagccca gatctgtcat ctggtgctgc cagctgtgct cactctggtt ttctgctcag 1320
ggtctgaagc agctgctgtc tccctcctct gccccatcc cctggctctc ccctggcac 1380
agtgccactc ccttggaagg gagggaacca cccgtgagcc ccaggggcttg ggaagcctga 1440
ggcgggcctc tgcctctccc tgccccagc acaattggca gagatgaggc gggtggtgga 1500
cggctgggct gtcgtggcag ggtctgcaca gggccatgtc ctggctgtaa cccaggcagt 1560
gggaggtctg cagcctggtc atggcctcca cagcaggtcc ctgtggacag acatatctgc 1620
```

atatttatca ataaagcctt ttgctccttc aaaaaaaaaa a 1661

<210> 46
<211> 1910
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2323917CB1

<400> 46
gtactccgcg cgcagccctg gcggtccagc gcgaatctcc agaatctggt ctccagaaac 60
attatttcaa tctgtgtatg gaaaaatccc agaaaattaa tccttttata ttgcatatac 120
tccaagaagt ggatgaagaa attaaaaagg ggctagcagc aggaatcaca ttaaacattg 180
ctggtaacaa tcggctagtg ccagtagaaa gagttacagg tgaagatttt tggattcttt 240
ccaaaatttt aaagaattgt ctgtatatta atggtttgga tgttggatat aaccttctat 300
gtgatgttgg tgcatactat gctgcgaaac tgcttcagaa acaacttaat ctcatttact 360
taaacctcat gtttaatgat attgggccag aaggtggaga attgattgct aaagtgctac 420
ataagaatcg gactctgaaa tacctaagaa tgactggaaa caaaattgaa aataaaggtg 480
gaatgttttt tgctgcaatg ctgcaaatta attcatcatt agagaaatta gatctgggtg 540
actgtgatct gggaatgcaa agtgtgatag catttgctac agtactaact caaaaccaag 600
caattaaggc aataaaccta aaccgaccta tactgtacgg tgaaacaggaa gagtctacag 660
tccatgtagg cctcatgttg aaagaaaatc actgtcttgt tgcactacac atgtgtaagc 720
atgatataaa aaacagtggt atacaacagt tatgtgatgc actgtatctg aacagtagcc 780
tgcgctacct tgatgtcagc tgcaacaaaa taactcatga tggaatggtg tatttggctg 840
atgtactgaa aagcaacact accctggaag taatagatct ttcctttaac agaatagaaa 900
atgcaggcgc aaactatctc agtgaaactc ttacttcaca caacagcagt cttaaagcgt 960
tgtcagtagt cagcaacaac atagagggag aaggacttgt tgcactttca caatcaatga 1020
aaacaaatct cactttctct catatctaca tttgggggaaa caaatttgat gaggctacgt 1080
gtatagcata ttcagactta attcaaatgg gttgtctaaa accagacaat acagatgtgg 1140
agccattttgt ggtagatgga cgtgtatatc ttgcagaagt ctccaatggc cttaaaaagc 1200
attattattg gacatcaact tatggagaat cttatgacca ctcatctaat gcaggttttg 1260
ctcttgttcc agtaggtcaa cagccatgaa aaagtagctg taaaataatt ttcatacact 1320
tgtcttattg tttcacagag attaatattc tatagcctat tttcttttat aaattaaaac 1380
aagttacttt tttcaaatgt gtaaaagata atgactgtat taaactttgt ataactgtca 1440
aaaaaaaaaa aaaggggggg gccccgatta gtgagctcgt ccccgcgggt atttactccc 1500
ggggcccggt cccttgggag gggttaacca gtctttccgc caagggtggt gtcctgtata 1560
tacaggcgtc ttggaaagag aagtcgtgct gtccctgggc tcttcttctg gtcggagggc 1620
attttactcg tctgtccccg ctggtgtaat atagtaccgc cacgcgtgtg gtccgtggta 1680
tgttggaaat aagtgcttgg tggagtctcg ctcggtgttg gacgtaggta gatcctcgca 1740
gacgttctca tatcggcttg gcgttggtcg tggcccgct tcccgtggtt gtgttgtcga 1800
ggtagtggtg ttgcacctgt atttatattg gtcctccgc cgcgcggggt ggaggcgtgg 1860
ttggttagt agctcgtggc ggttgttgtg ttcctcaacc cacacacaat 1910

<210> 47
<211> 2162
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2754960CB1

<400> 47
gacagcgcaa ccgcactcca gcctggacga cagaatgaga ctccgtctca aaaaaaaaaa 60
aaaaaaaaaa aagcgagggc tggaagtctg gagcaggtgc gcggctgcaa cggcagccgc 120
gggaagctcg ggccggcagg gtttccccgc acgctggcgc ccagctccgg gcgcggagcc 180
cgctgtaagt ttcgctttcc attcagtgga aaacgaaagc tgggcggggt gccacgagcg 240
cggggccaga ccaaggcggg cccggagcgg aacttcggtc ccagctcggt cccccggctca 300
gtcccgacgt ggaactcagc agcggaggct ggacgcttgc atggcgcttg agagattcca 360
tcgtgcctgg ctcacataag cgcttcctgg aagtgaagtc gtgctgtcct gaacgcgggc 420
caggcagctg cggcctgggg gttttggagt gatcacgaat gagcaaggcg tttgggctcc 480
tgaggcaaat ctgtcagtcc atcctggctg agtcctcgca gtccccggca gatcttgaag 540
aaaagaagga agaagacagc aacatgaaga gagagcagcc cagagagcgt cccaggggcct 600
gggactaccc tcatggcctg gttggtttac acaacattgg acagacctgc tgccttaact 660
ccttgattca ggtgttcgta atgaatgtgg acttcaccag gatattgaag aggatcacgg 720

```
tgcccagggg agctgacgag cagaggagaa gcgtcccttt ccagatgctt ctgctgctgg 780
agaagatgca ggacagccgg cagaaagcag tgcggcccct ggagctggcc tactgcctgc 840
agaagtgcaa cgtgcccttg tttgtccaac atgatgctgc ccaactgtac ctcaaactct 900
ggaacctgat taaggaccag atcactgatg tgcacttggt ggagagactg caggccctgt 960
atacgatccg ggtgaaggac tccttgattt gcgttgactg tgccatggag agtagcagaa 1020
acagcagcat gctcaccctc ccacttttctc tttttgatgt ggactcaaag cccctgaaga 1080
cactggagga cgcctgcac tgcttcttcc agcccaggga gttatcaagc aaaagcaagt 1140
gcttctgtga gaactgtggg aagaagaccc gtgggaaaca ggtcttgaag ctgacccatt 1200
tgccccagac cctgacaatc cacctcatgc gattctccat caggaattca cagacgagaa 1260
agatctgcca ctccctgtac ttcccccaga gcttggattt cagccagatc cttccaatga 1320
agcgagagtc ttgtgaggct gaggagcagt ctggagggca gtatgagctt tttgctgtga 1380
ttgcgcacgt gggaatggca gactccggtc attactgtgt ctacatccgg aatgctgtgg 1440
atggaaaatg gttctgcttc aatgactcca atatttgctt ggtgtcctgg gaagacatcc 1500
agtgtaccta cggaaatcct aactaccact ggcaggaaac tgcatatctt ctggtttaca 1560
tgaagatgga gtgctaatgg aaatgcccaa aaccttcaga gattgacacg ctgtcatttt 1620
ccatttccgt tcctggatct acggagtctt ctaagagatt ttgcaatgag gagaagcatt 1680
gttttcaaac tatataactg agccttattt ataattaggg atattatcaa aatatgtaac 1740
catgaggccc ctcaggtcct gatcagtcag aatggatgct ttcaccagca gacccggcca 1800
tgtggctgct cggtcctggg tgctcgctgc tgtgcaagac attagccctt tagttatgag 1860
cctgtgggaa cttcaggggt tcccagtggg gagagcagtg gcagtgggag gcatctgggg 1920
gccaaaggtc agtggcaggg ggtatttcag tattatacaa ctgctgtgac cagacttgta 1980
tactggctga atatcagtgc tgtttgtaat ttttcacttt gagaaccaac attaattcca 2040
tatgaatcaa gtgttttgta actgctattc atttattcag caaatattta ttgatcatct 2100
cttctccata agatagtgtg ataaacacag tcatgaataa agttattttc cacaaaaaaa 2160
aa                                                              2162

<210> 48
<211> 578
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3092341CB1

<400> 48
gaaaggccca cctgtgtcct ggttgagggt ctccagggtt ctttggggcc cgaggaattt 60
cagacctgaa caatcaggtg gacaacttct tgtgactttg ggtattctga aattactcat 120
cttgactcta ctgtgcctgg atgtagctgt gtgttttgtg tgtatcctgt tcatggccaa 180
tggtggcaga gtctacatag aactatgctt cgtggtgttc tggggaaaac ctttcgactt 240
gttggctata ctattcaata tggctgtata gctcattgtg cttttgaata cgttggtggt 300
gttgtcatgg tgccaatggg tcatgtttgg ttagaaggtg acaatctaca gaattctaca 360
gattccaggt gctatggacc tattccatat ggactaataa gaggacgaat cttctttaag 420
atttggcttc tgagtgattt tggattttta cgtgccagcc ctaatggcca cagatttct 480
gatgattagt aagcatttat tcttttgact tgattattgt ctccttttca tgtgaattta 540
ttactcccgt tgaaaccgtg tacttaccaa taaactat                        578

<210> 49
<211> 1300
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3658034CB1

<400> 49
agggaatccc aattatgagg cggctaagca ggcaaagaaa tagcttgggt ggttttgcat 60
ctactcagga taccactgtg gctttaaagg ctgtctgaat ttgcagccct aatgaatcca 120
gaaaggacaa atatccaagt gaccgtgacg gggcctagct caccaagtcc tgtaaagttt 180
ctgattgaca cacacaaccg cttactcctt cagacagcag agcttgctgt ggtacagcca 240
acggcagtta atatttccgc aaatggtttt ggatttgcta tttgtcagct caatgttgta 300
tataatgtga aggctttctgg gtcttctaga agacgaagat ctatccaaaa tcaagaagcc 360
tttgatttag atgttgctgt aaaagaaaat aaagatgatc tcaatcatgt ggatttgaat 420
gtgtgtacaa gcttttcggg cccgggtagg agtggcatgg ctcttatgga agttaaccta 480
ttaagtggct ttatggtgcc ttcagaagca atttctctga gcgagacagt gaagaaagtg 540
gaatatgatc atggaaaact caacctctat ttagattctg taaatgaaac ccagttttgt 600
```

gttaatattc ctgctgtgag aaactttaaa gtttcaaata cccaagatgc ttcagtgtcc 660
atagtggatt actatgagcc aaggagacag gcggtgagaa gttacaactc tgaagtgaag 720
ctgtcctcct gtgaccttg cagtgatgtc cagggctgcc gtccttgtga ggatggagct 780
tcaggctccc atcatcactc ttcagtcatt tttattttct gtttcaagct tctgtacttt 840
atggaacttt ggctgtgatt tatttttaaa ggactctgtg taacactaac atttccagta 900
gtcacatgtg attgttttgt tttcgtagaa gaatactgct tctattttga aaaaagagtt 960
ttttttcttt ctatgggggtt gcagggatgg tgtacaacag gtcctagcat gtatagctgc 1020
atagatttct tcacctgatc tttgtgtgga agatcagaat gaatgcagtt gtgtgtctat 1080
atttcccct ctcaaaaatct tttagaattt ttttggaggt gtttgttttc tccagaataa 1140
aggtattact ttagaatagg tattctcctc attttgtgaa agaaatgaac ctagattctt 1200
aagcattatt acacatccat gtttgcttaa agatggattt ccctgggaat gggagaaaac 1260
agccagcagg aggagcttca tctgttccct tcccacctcc 1300

<210> 50
<211> 2241
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3883861CB1

<400> 50
gctccccgag ccgggctgca ccggaggcgg cgagatggtc gcgcgcgtcg gcctcctgct 60
gcgcgccctg cagctgctac tgtgggggca cctggacgcc cagcccgcgg agcgcggagg 120
ccaggagctg cgcaaggagg cggaggcatt cctagagaag tacggatacc tcaatgaaca 180
ggtccccaaa gctcccacct ccactcgatt cagcgatgcc atcagagcgt ttcagtgggt 240
gtcccagcta cctgtcagcg gcgtgttgga ccgcgccacc ctgcgccaga tgactcgtcc 300
ccgctgcggg gttacagata ccaacagtta tgcggcctgg gctgagagga tcagtgactt 360
gtttgctaga caccggacca aaatgaggcg taagaaacgc tttgcaaagc aaggtaacaa 420
atggtacaag cagcacctct cctaccgcct ggtgaactgg cctgagcatc tgccggagcc 480
ggcagttcgg ggcgccgtgc gcgccgcctt ccagttgtgg agcaacgtct cagcgctgga 540
gtttctgggag gccccagcca caggccccgc tgacatccgg ctcaccttct tccaagggga 600
ccacaacgat gggctgggca atgccttgga tggcccaggg ggcgccctgg cgcacgcctt 660
cctgccccgc cgcggcgaag cgcacttcga ccaagatgag cgctggtccc tgagccgccg 720
ccgcgggcgc aacctgttcg tggtgctggc gcacgagatc ggtcacacgc ttggcctcac 780
ccactcgccc gcgccgcgcg cgctcatggc gccctactac aagaggctgg gccgcgacgc 840
gctgctcagc tgggacgacg tgctggccgt gcagagcctg tatgggaagc ccctaggggg 900
ctcagtggcc gtccagctcc caggaaagct gttcactgac tttgagacct gggactccta 960
caggccccaa ggaaggcgcc ctgaaacgca gggccctaaa tactgccact cttccttcga 1020
tgccatcact gtagacaggc aacagcaact gtacattttt aaagggagcc atttctggga 1080
ggtggcagct gatggcaacg tctcagagcc ccgtccactg caggaaagat gggtcgggct 1140
gcccccaac attgaggctg cggcagtgtc attgaatgat ggagatttct acttcttcaa 1200
aggggtcga tgctggaggt tccgggggcc caagccagtg tgggtctcc cacagctgtg 1260
ccgggcaggg ggcctgcccc gccatcctga cgccgccctc ttcttcctc ctctgcgccg 1320
cctcatcctc ttcaagggtg cccgctacta cgtgctggcc cgaggggggac tgcaagtgga 1380
gccctactac ccccgaagtc tgcaggactg gggaggcatc cctgaggagg tcagcggcgc 1440
cctgccgagg cccgatggct ccatcatctt cttccgagat gaccgctact ggcgcctcga 1500
ccaggccaaa ctgcaggcaa ccacctcggg ccgctgggcc accgagctgc cctggatggg 1560
ctgctggcat gccaactcgg ggagcgccct gttctgaagg cacctcctca cctcagaaac 1620
tggtggtgct ctcagggcaa aatcatgttc cccacccccg gggcagaacc cctcttagaa 1680
gcctctgagt ccctctgcag aagaccgggc agcaaagcct ccatctggaa gtctgtctgc 1740
ctttgttcct tgaagaatgc agcattgtct ttgtctgtcc ccaccacatg gaggtggggg 1800
tgggatcaat cttaggaaaa gcaaaaaagg gtcccagatc ccttggccct ttcctccgag 1860
gacttctatc ctccccaggc ctttgtttct tcggctaaag gtacagttcc tttcaagagg 1920
taacagcact gggatccaag caggggatg aaaaactcag cagagaaatt cgagaccatt 1980
ttgcaagact gtgcccttct cctcaggacc ccctggctca gttcttgaaa aacggtgtca 2040
tatttagtca gaggccccac ccccaggaag catggatggg gatgaaggca caggcgtctc 2100
caaccтcaga ggccctttgt ggggtcagga cacagagtgg gagggagact gatgcaggcc 2160
taccagtccc tggcttttg tctggggctg gaataaagag gtgccttcag ctggtgggcc 2220
gagaggcagg aaaaaaaaaa a 2241

<210> 51
<211> 1860
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4993873CB1

<400> 51
agaattgacc tgacaaatat tgaccaggtt gctgtgatct tcaaacacca ttttcctgtg 60
ggtagaggtg atgcagtgct gaagacctgg gccctgctc agtgcctttg ctctagaatg 120
ggtccagctt ggctttggct actgggaaca gggatcctgg cctctgtcca ctgtcagccc 180
cttcttgccc atggagataa aagtctgcag gggcctcaac ccccaggca tcagctctca 240
gagccagccc ccgctacaca cagaatcaca cccaccatta ccaattttgc tttgcgtttg 300
tataaagagc tggcagcaga cgcccccgga aacatcttct tctcgccagt gagcatctcc 360
accaccctgg cctgctctc tcttgggggcc caagctaaca cctcagctct gatcctggag 420
ggcctgggat tcaacctcac agaaaccct gaagccgaca tccaccaggg cttccggagc 480
ctcctccaca ccctgccct gcccagcccc aaactcgaac taaaagtagg aaactcctg 540
ttcctagaca agcgactaaa gcctcggcag cactatttgg acagcatcaa ggagctttat 600
ggagcttttg ctttttctgc caacttcaca gattctgtta caactgggag gcagattaat 660
gactatttga gaaggcaaac atacgggcaa gtcgtggact gcctcccgga gttcagccag 720
gacacgttca tggttcttgc caattacatc ttcttcaaag ccaagtggaa gcaccctttc 780
agtcgctacc agacccagaa gcaggcaagt ttctttgtgg atgagaggac ttctctccag 840
gtccccatga tgcaccaaaa ggaaatgcac agattcctct atgaccagga tttggcttgc 900
actgtcctcc agatagaata cagaggaaat gccttggcgc tgctggtcct ccctgaccg 960
gggaaaatga agcaggtgga ggctgctctg cagccacaga ccctgagaaa atgtgggccaa 1020
ttgctcctgc ccagtctgtt ggatttgcac ttgccaaggt tttcaatttc tggaacatat 1080
aacctggaag acatacttcc ccaaattggt ctcaccaaca tactcaactt agaagctgac 1140
ttctcaggag tcactgggca gctcaacaaa accatctcca aggtgtcaca caaggcgatg 1200
gtggacatga gtgagaaggg gaccgaggcc ggggctgctt caggcctcct ctcccagccc 1260
ccatctctga acaccatgtc agacccacat gcccacttca acaggccttt cctcttgctc 1320
ctttgggagg tcaccaccca gagcttactc ttcctgggaa aagttgtcaa cccagttgca 1380
gggtaaccat ggtgggaggc caggagttat cttatctcat cctggaccaa acagataggc 1440
cagaaccagc ctgcatcctg gggctgctat gtggttcagt taatcagtgt ggccaagattc 1500
taataaagtt gaccttgggt tctgtgaaaa aaaaaaaaag aacaaccaac gcataaagtt 1560
gcggcgtaaa ggatgctgaa agcaaatacc ccggtttgcc cgtccaaatt taacataagg 1620
gggggccgg cagtacagaa ggtccctcag agtgtgtttg atggccgtcg actgattggc 1680
gggggggcgg agagagtttt tacgcatctt tgcacgtgtt agcacccggg gggggggggc 1740
ccataggacc cccggagctg gagggttgtt ttgcgccggc actcttgggg gggtgtttat 1800
gaccgacttt tagtggcaca aactgtgggg cttctcatat gtgtgcgcgc gcttctgttt 1860

<210> 52
<211> 550
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5208004CB1

<400> 52
ggcgggtggc tggaggacac ggtgtaggcc ttgccgctgt ctgggttcct cgccactgca 60
ggagcaggac tgtttctgga aacactgggc tgctggtgac ccgtcaaact tctcccacaa 120
attctgaatc cgagaaagtg aaggaaggat ggtgggggaag tgaggaggca ggagcagagg 180
ccacagggac cgaccagaga tgcggtggag acagaggagc ttcctctcta ggctgtttct 240
gggaagcctg agaggcggcc agcaccacc tccgctcact ctcccctcag cctcttcgct 300
tcccttctca acccttctc tcctcctcgc ctcttccctt tcttgctgtc ttgttagtcc 360
ctgtcccaaa actcctggct cctttgttct gctgccgtgg cccccaccca ggaggaggtc 420
tcaggcacca tccccccac gagggatcca cacaacaggt tcatgctggg gctgggggag 480
ccccgctggg ttcctgatgc cttgtgcaca gggaagtgct gcagtcattt ttggactctc 540
ctgaatgtgt 550

<210> 53
<211> 2597
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5267783CB1

<400> 53
gcctggttgg gccgcgtcct tctgctgccc cgcgccggtg caggcctcgc cgcgagccgc 60
aggtgtcctg gagtctggcc caggacctgg ccccacagga gtcccagcag gggtagctcc 120
tcccgggaca aggaccgaag tgcgacggtc agtagttcag tgcccatgcc tgctggaggg 180
aaaggaagcc atccttcatc tacaccccag agggtcccca accgcctgat ccacgagaag 240
tcaccatacc tcctacaaca tgcctacaat cctgtggact ggtacccctg gggagaggaa 300
gccttcgaca aggccaggaa ggaaaacaag ccgattttcc tctcagtcgg gtactccacc 360
tgccactggt gccacatgat ggaagaggag tccttccaga atgaggagat tggccgcctg 420
ctcagtgagg actttgtgag tgtgaaggta gaccgtgagg agcggcctga cgtggacaag 480
gtgtacatga cgttcgtgca ggccaccagc agcggcgggg gctggcccat gaatgtgtgg 540
ctgactccca acctccagcc ctttgtcggg ggcacctatt tccctcctga ggatggcttg 600
acccgagtcg gcttccgcac agtgttgctg agaatacgag aacagtggaa acagaacaag 660
aacaccctgc tagaaaatag ccagcgtgtc accactgccc tgctggcccg atcagagatc 720
agcgtgggtg accgccagct gccgccctct gccgccaccg tgaacaatcg ctgcttccag 780
cagctggatg agggctatga tgaggaatac ggtggcttcg ctgaggcccc caagtttccc 840
acgccggtga tcctgagctt cctgttctcc tactggctca gccatcgact gactcaggat 900
ggctctcggg cccagcagat ggccttgcat accctgaaaa tgatggctaa cgggggcatc 960
cgggaccatg tggggcaggg ctttcaccgc tactccacag accgccagtg gcacgtccct 1020
cagctctctg gtgatgaatt ctactctgac gtggccaaag gcatcctgca gtacgtggct 1140
cggagcctga gccaccggtc cggaggcttc tatagcgcag aagatgcaga ctcgccccca 1200
gagcggggcc agcggcccaa agagggcgcc tactatgtgt ggacggtcaa agaggttcag 1260
cagctcctcc cggagcctgt gttgggtgcc accgagccgc tgacctcagg ccagctcctc 1320
atgaagcact acggcctcac agaggctggt aacatcagcc ccagtcagga cccacaaggg 1380
gagctgcagg gccagaatgt gctgaccgtc cggtactcgc tggagctgac tgctgcccgc 1440
tttggcttgg atgtgggagc cgtgcggacc ttgctcaatt cagggctgga gaagctcttc 1500
caggccggga agcatcggcc caagccgcac ctggacagca agatgctggc tgcctggaat 1560
ggcttgatgg tgtcaggcta tgctgtgact ggggctgtcc tgggccaaga caggctgatc 1620
aactatgcca ccaatggtgc caagttcctg aagcggcaca tgtttgatgt ggccagtggc 1680
cgcctgatgc ggacctgcta caccggccct gggggactg tggagcacag caacccaccc 1740
tgctggggct tcctggagga ctacgccttc gtggtgcggg gcctgctgga cctgtatgag 1800
gcctcacagg agagtgcgtg gctcgagtgg gctctgcggc tgcaggacac acaggacagg 1860
ctcttttggg actcccaggg tggcggctac ttctgcagtg aggctgagct gggggctggc 1920
ctgcccctgc gtctgaagga cgaccaggat ggagcagagc ccagcgccaa ttccgtgtca 1980
gcccacaacc tgctccggct gcatggcttc acgggccaca aggactggat ggacaagtgt 2040
gtgtgcctat tgaccgcctt ttccgagcgc atgcgtcgtg tcccggtggc gttgcccgag 2100
atggtccgcg cctctcagc ccagcagcag accctcaagc agatcgtgat ctgtggagac 2160
cgtcaggcca aggacaccaa ggccctggtg cagtgcgtcc actctgtcta cattcctaac 2220
aaggtgctga ttctggctga tgggaccccc tcgagcttcc tgtcccgcca gctgcctttc 2280
ctgagtaccc tccgacggtt ggaagaccag gccactgcat atgtgtgtga gaatcaagcc 2340
tgctcagtgc ccatcactga tccctgcgaa ttacgaaaac tactacatcc atgactgccc 2400
caacccccct gggtggggc agaaggtgaa gcatcccaac tgactagaga ctcaggccct 2460
gcagggccct atagaacctg tggccatccc tgagcaccct gccaccaggt gacctcggcc 2520
atactcactg ccccccttgg gcacccactc accctagaat aaacttaaca gtgtcccgtg 2580
gtaaaaaaaa aaaaaaa 2597

<210> 54
<211> 2385
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5583922CB1

<400> 54
gaccgacccg ccacccgccg tagcccgcgc gcccctggca ctcaatcccc gccatgtggg 60
ggctcctgct cgcctggcc gccttcgcgc cggccgtcgg cccggctctg ggggcgccca 120
ggaactcggt gctgggcctc gcgcagcccg ggcaccacaa ggtcccaggc tcgacccccg 180
ccctgcatag cagcccggca cagccgccgg cggagacagc taacgggacc tcagaacagc 240
atgtccggat tcgagtcatc aagaagaaaa aggtcattat gaagaagcgg aagaagctaa 300
ctctaactcg ccccacccca ctggtgactg ccgggcccct tgtgacccca actccagcag 360
ggaccctcga ccccgctgag aaacaagaaa caggctgtcc tcctttgggt ctggagtcct 420
tgcgagtttc agatagccgg cttgaggcat ccagcagcca gtccttggt cttggaccac 480
accgaggacg gctcaacatt cagtcaggcc tggaggacgg cgatctatat gatggagcct 540
ggtgtgctga ggagcaggac gccgatccat ggtttcaggt ggacgctggg caccccaccc 600
gcttctcggg tgttatcaca cagggcagta actctgtctg gaggtatgac tgggtcacat 660

```
catacaaggt ccagttcagc aatgacagtc ggacctggtg gggaagtagg aaccacagca 720
gtgggatgga cgcagtattt cctgccaatt cagacccaga aactccagtg ctgaacctcc 780
tgccggagcc ccaggtggcc cgcttcattc gcctgctgcc ccagacctgg ctccagggag 840
gcgcgccttg cctccgggca gagatcctgg cctgcccagt ctcagacccc aatgacctat 900
tccttgaggc ccctgcgtcg ggatcctctg accctctaga ctttcagcat cacaattaca 960
aggccatgag gaagctgatg aagcaggtac aagagcaatg ccccaacatc acccgcatct 1020
acagcattgg gaagagctac cagggcctga agctgtatgt gatggaaatg tcggacaagc 1080
ctggggagca tgagctgggg gagcctgagg tgcgctacgt ggctggcatg catgggaacg 1140
aggccctggg gcgggagttg cttctgctcc tgatgcagtt cctgtgccat gagttcctgc 1200
gagggaaccc acgggtgacc cggctgctct ctgagatgcg cattcacctg ctgccctcca 1260
tgaaccctga tggctatgag atcgcctacc accggggttc agagctggtg ggctgggccg 1320
agggccgctg gaacaaccag agcatcgatc ttaaccataa ttttgctgac ctcaacacac 1380
cactgtggga agcacaggac gatgggaagg tgccccacat cgtccccaac catcacctgc 1440
cattgcccac ttactacacc ctgccaatg ccaccgtggc tcctgaaacg cgggcagtaa 1500
tcaagtggat gaagcggatc ccctttgtgc taagtgccaa cctccacggg ggtgagctcg 1560
tggtgtccta cccattcgac atgactcgca ccccgtgggc tgcccgcgag ctcacgccca 1620
caccagatga tgctgtgttt cgctggctca gcactgtcta tgctggcagt aatctggcca 1680
tgcaggacac cagccgccga ccctgccaca gccaggactt ctccgtgcac ggcaacatca 1740
tcaacggggc tgactggcac acggtccccg ggagcatgaa tgacttcagc tacctacaca 1800
ccaactgctt tgaggtcact gtggagctgt cctgtgacaa gttcctcac gagaatgaat 1860
tgccccagga gtgggagaac aacaaagacg ccctcctcac ctacctggag caggtgcgca 1920
tgggcattgc aggagtggtg agggacaagg acacggagct tgggatgct gacgctgtca 1980
ttgccgtgga tgggattaac catgacgtga ccacggcgtg gggcgggat tattggcgtc 2040
tgctgacccc aggggactac atggtgactg ccagtgccga gggctaccat tcagtgacac 2100
ggaactgtcg ggtcaccttt gaagagggcc ccttccctg caatttcgtg ctcaccaaga 2160
ctcccaaaca gaggctgcgc gagctgctgg cagctggggc caaggtgccc ccggacctc 2220
gcaggcgcct ggagcggcta aggggacaga aggattgata cctgcggttt aagagcccta 2280
gggcaggctg gacctgtcaa gacgggaagg ggaagagtag agaggagggg acaaagtgag 2340
gaaaaggtgc tcattaaagc taccgggcac cttaaaaaaa aaaaa 2385
```

Docket No.: PF-0727 USN

DECLARATION AND POWER OF ATTORNEY FOR UNITED STATES PATENT APPLICATION

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name, and

I believe that I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if more than one name is listed below) of the subject matter which is claimed and for which a United States patent is sought on the invention entitled

PROTEASES AND PROTEASE INHIBITORS

the specification of which:

/___/ is attached hereto.

/___/ was filed on _____ as application Serial No. _____ and if this box contains an X /___/, was amended on _____.

/X__/ was filed as Patent Cooperation Treaty international application No. PCT/US00/21878 on August 9, 2000, if this box contains an X /_/, was amended on under Patent Cooperation Treaty Article 19 on _____ 2001, and if this box contains an X /_/, was amended on _____.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge my duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56(a).

I hereby claim the benefit under Title 35, United States Code, §119 or §365(a)-(b) of any foreign application(s) for patent or inventor's certificate indicated below and of any Patent Cooperation Treaty international applications(s) designating at least one country other than the United States indicated below and have also identified below any foreign application(s) for patent or inventor's certificate and Patent Cooperation Treaty international application(s) designating at least one country other than the United States for the same subject matter and having a filing date before that of the application for said subject matter the priority of which is claimed:

Country	Number	Filing Date	Priority Claimed
_____	_____	_____	/_/ Yes /_/ No
_____	_____	_____	/_/ Yes /_/ No

I hereby claim the benefit under Title 35, United States Code, §119(e) of any United States provisional application(s) listed below.

Application Serial No.	Filed	Status (Pending, Abandoned, Patented)
60/147,986	August 9, 1999	Expired
60/160,807	October 21, 1999	Expired

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in said prior application(s) in the manner required by the first paragraph of Title 35, United States Code §112, I acknowledge my duty to disclose material information as defined in Title 37 Code of Federal Regulations, §1.56(a) which occurred between the filing date(s) of the prior application(s) and the national or Patent Cooperation Treaty international filing date of this application:

Application Serial No.	Filed	Status (Pending, Abandoned, Patented)

I hereby appoint the following:

Lucy J. Billings	Reg. No. 36,749
Michael C. Cerrone	Reg. No. 39,132
Diana Hamlet-Cox	Reg. No. 33,302
Richard C. Ekstrom	Reg. No. 37,027
Barrie D. Greene	Reg. No. 46,740
Lynn E. Murry	Reg. No. 42,918
Shirley A. Recipon	Reg. No. 47,016
Susan K. Sather	Reg. No. 44,316
Michelle M. Stempien	Reg. No. 41,327
David G. Streeter	Reg. No. 43,168
Stephen Todd	Reg. No. 47,139
P. Ben Wang	Reg. No. 41,420

respectively and individually, as my patent attorneys and/or agents, with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith. Please address all communications to:

LEGAL DEPARTMENT
INCYTE GENOMICS, INC.
3160 PORTER DRIVE, PALO ALTO, CA 94304

TEL: 650-855-0555 FAX: 650-849-8886 or 650-845-4166

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

1-00

First Joint Inventor:

Full name: Henry Yue

Signature: Henry Yue

Date: September 24, 2001

Citizenship: United States

Residence: Sunnyvale, California CA

P.O. Address: 826 Lois Avenue
Sunnyvale, California 94087

2-00

Second Joint Inventor:

Full name: Preeti Lal

Signature: Preeti Lal

Date: September 10, 2001

Citizenship: India

Residence: Santa Clara, California CA

P.O. Address: P.O. Box 5142
Santa Clara, California 95056

3-00

Third Joint Inventor:

Full name: Y. Tom Tang

Signature:

Date: Sept. 10, 2001

Citizenship: United States

Residence: San Jose, California CA

P.O. Address: 4230 Ranwick Court
San Jose, California 95118

4-00

Fourth Joint Inventor:

Full name: Olga Bandman

Signature: Olga Bandman

Date: 12 September, 2001

Citizenship: United States

Residence: Mountain View, California CA

P.O. Address: 366 Anna Avenue
Mountain View, California 94043

5-00

Fifth Joint Inventor:

Full name: Mariah R. Baughn

Signature: Mariah R. Baughn

Date: September 5, 2001

Citizenship: United States

Residence: San Leandro, California CA

P.O. Address: 14244 Santiago Road
San Leandro, California 94577

6-00

Sixth Joint Inventor:

Full name: Yalda Azimzai

Signature:

Date: September 13, 2001

Citizenship: United States

Residence: Castro Valley, California CA

P.O. Address: 5518 Boulder Canyon Drive
Castro Valley, California 94552

7-00

Seventh Joint Inventor:

Full name: Dyung Aina M. Lu

Signature:

Date: Sept 7, 2001

Citizenship: United States

Residence: San Jose, California CA

P.O. Address: 233 Coy Drive
San Jose, California 95123

8-00

Eighth Joint Inventor:

Full name: Junming Yang

Signature:

Date: September 17, 2001

Citizenship: China

Residence: San Jose, California CA

P.O. Address: 7125 Bark Lane
San Jose, California 95129